

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 08:44:48 ; Search time 2017.73 seconds  
(without alignments)  
5714.937 Million cell updates/sec

Title: US-10-008-355-2  
Perfect score: 3719  
Sequence: 1 MQMKLSILLGAALLGASG.....LFMDKWKQCPRLIQELKLI 712

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US10008355/runat\_16052003.110401\_2551/app\_query.fasta\_1.1230  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355 -ECGN.1.1.2013 -runat\_16052003.110401\_2551 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_fod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	418.5	11.3	591	17	BH392840	BH392840 AG-ND-162
C 2	397	10.7	500	17	BH386991	BH386991 AG-ND-148
C 3	397	10.7	537	17	BH388186	BH388186 AG-ND-124
C 4	385	10.4	418	17	BH375201	BH375201 AG-ND-173
C 5	315	8.5	697	17	BH375688	BH375688 AG-ND-120
C 6	314	8.4	426	17	BH400391	BH400391 AG-ND-147
C 7	239.5	6.4	398	10	AW774807	AW774807 EST33958
C 8	187.5	5.0	721	17	BH391233	BH391233 AG-ND-141
C 9	184	4.9	706	17	BH394089	BH394089 AG-ND-150
C 10	177	4.8	555	17	BH400866	BH400866 AG-ND-125
C 11	177	4.8	769	17	BH381877	BH381877 AG-ND-155
C 12	160.5	4.3	1005	17	BH164253	BH164253 ENFTG63TF
C 13	138	3.7	262	17	BH391877	BH391877 AG-ND-141
C 14	136	3.7	822	17	BH371846	BH371846 AG-ND-162
C 15	116.5	3.1	1869	11	AY107109	AY107109 Zea mays
C 16	107.5	2.9	1016	14	BM800126	BM800126 AGENCOURT
C 17	107.5	2.9	2349	11	AK012962	AK012962 Mus muscu
C 18	107.5	2.9	3112	11	AK004749	AK004749 Mus muscu
C 19	103.5	2.8	1557	17	BH770729	BH770729 LLMGtaq47
C 20	103	2.8	868	10	AW057494	AW057494 ca04f04.x
C 21	102	2.7	663	10	AW057482	AW057482 ca04c04.x
C 22	102	2.7	864	9	AU205006	AU205006 AU205006
C 23	102	2.7	1095	11	BC015158	BC015158 Homo sapi
C 24	101.5	2.7	1012	9	AL545921	AL545921 AL545921
C 25	101.5	2.7	3340	11	AK004800	AK004800 Mus muscu
C 26	101.5	2.7	3369	11	AK011711	AK011711 Mus muscu
C 27	100.5	2.7	651	10	BE266416	BE266416 601193163
C 28	100.5	2.7	2720	11	AY104793	AY104793 Zea mays
C 29	100	2.7	937	12	BF105351	BF105351 601822064
C 30	99.5	2.7	613	14	BQ981366	BQ981366 QGE13J08.
C 31	99	2.7	1265	17	BH770553	BH770553 LLMGtaq31
C 32	98.5	2.6	1131	17	CNS061E4	AL404066 T3 end of
C 33	98	2.6	667	13	BI391728	BI391728 pppln.pk0
C 34	98	2.6	717	9	AL717307	AL717307 AL717307
C 35	98	2.6	723	9	AL717257	AL717257 AL717257
C 36	98	2.6	730	9	AL726954	AL726954 AL726954
C 37	98	2.6	1087	12	BE799042	BE799042 601583645
C 38	97.5	2.6	920	9	AL545962	AL545962 AL545962
C 39	97.5	2.6	957	17	AZ680449	AZ680449 ENTHN69TR
C 40	97.5	2.6	2836	11	AK004712	AK004712 Mus muscu
C 41	97	2.6	720	9	AL720110	AL720110 AL720110
C 42	97	2.6	946	12	BF685895	BF685895 602143193
C 43	96.5	2.6	713	14	BU004296	BU004296 OGG4b07.y
C 44	96.5	2.6	2232	11	BC016514	BC016514 Homo sapi
C 45	96	2.6	1063	14	BM909800	BM909800 AGENCOURT

ALIGNMENTS

RESULT 1  
BH392840/c  
LOCUS BH392840  
DEFINITION AG-ND-162P14.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162P14  
DNA sequence.  
591 bp DNA linear GSS 11-DEC-2001  
ACCESSION BH392840  
VERSION BH392840.1 GI:17338981  
KEYWORDS GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
REFERENCE 1 (bases 1 to 591)  
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

Qy	262	TyrlYsAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu	281
Db	117	AATAAGCGCTGGTGATTTTGCAATGATTTCTAGGATATCTGGTACAAACAACAGATACGCTT	58
Qy	282	ThrSerTrpGlyValGluAspArgIleGluAsnGlu	293
Db	57	ACTTCTTCGGAATCGAACAAATCGTAGCAAAAGAT	22
RESULT 2			
LOCUS	BH386991/c		
DEFINITION	AG-ND-148L11.TR ND-TAM Anopheles gambiae genomic clone AG-ND-148L11		
ACCESSION	BH386991	500 bp	DNA
VERSION	BH386991		linear
KEYWORDS	BH386991.1		GSS 10-DEC-2001
SOURCE	GSS.		
ORGANISM	African malaria mosquito.		
	Anopheles gambiae		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;		

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1. (bases 1 to 500)  
Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.  
Direct Submission of BAC-end sequences from *Anopheles gambiae*  
Unpublished (2001)  
Other\_GSSs: AG-ND-148L11.TF  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjlloftus@igrr.org

This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center

```

University, College Station, Texas //843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES             Location/Qualifiers
     source            1..500
                        /organism="Anopheles gambiae"
                        /strain="PEST"
                        /db_xref="taxon:7165"
                        /clone="AG-ND-148L11"
                        /clone_lib="ND-TAM"
                        /notes="vector: pECBAC1; Site_1: HindIII"

BASE COUNT      143 a    105 c    90 g    162 t
ORIGIN

Alignment Scores:
Pred. ID.:          1.14e-39           Length:          500
Score:              397.00             Matches:         71
Percent Similarity: 65.00%             Conservative:    33
Best Local Similarity: 44.38%           Mismatches:     52
Query Match:        10.67%             Indels:          4
DB:                  17                 Gaps:            2

US-10-008-355-2 (1-712) x BH386991 (1-500)

Qy   134   LysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeu 153
      : : : : : | | | | | : : : : : : : : : : | | | | | : : : : :
Db   484   CGTATCACTTCTTAARCTGAACAATGATATGAGC-----CGTGATCAGAGAAAAGCTATT 431
      : : : : : | | | | | : : : : : : : : : : | | | | | : : : : :

Qy   154   ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGln 173
      : : : : : | | | | | : : : : : : : : : : | | | | | : : : : :
Db   430   ATCCGATCGTAAATCAAGGCTATCCAGCTCTCAAACACTCTCGAATGGA-----AAATAC 377
      : : : : : | | | | | : : : : : : : : : : | | | | | : : : : :
```

```

Qy 174 LeuCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspVal 193
Db 376 ACTGTAGTGGTGAAGGATTTCTTCAAGGAAATGAGTTTACTACTTTGTATTCAGCAT 317
Qy 194 PheLysAspValArgMetValPheAlaProSerSerValGlyLysPheGlyGlyAsp 213
Db 316 TTCAAAGATGTTCGTCTGGTACTCCGCGCATATTTGGTAAATATGGTGGTAT 257
Qy 214 ThrAspAsnTyrMetTyrProArgHisThrGlyAspPheSerValPheArgValTyrAla 233
Db 256 ACTGATACTGGGATGGCCAGACACACAGAGACTTCTCTGTTTCCGTTTATGCT 197
Qy 234 GlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyr 253
Db 196 GATAAAATGCAATCTCTGTAATATTCAGCTAACACGTTCCCTTTAAAGCCTAAGCAT 137
Qy 254 PheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhe 273
Db 136 CATTTACCAATTTCTACTAAAGGAAATTAAGCCTGGTATTTGCAATGATTGTAGGATAT 77
Qy 274 ProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGlu 293
Db 76 CTGTGTACAACAACAGATACCTTCTTTCGGAATCGAACAATCGTAAGCAACAT 17

RESULT 3
BH388186/c
LOCUS BH388186 537 bp DNA linear GSS 11-DEC-2001
DEFINITION AG-ND-124P3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-124P3,
DNA sequence.
ACCESSION BH388186
VERSION BH388186.1 GI:17334327
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 537)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-124P3.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
FEATURES
Location/Qualifiers
1..537
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-124P3"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 153 a 112 c 94 g 178 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 1.3e-39 Length: 537
Score: 397.00 Matches: 71
Percent Similarity: 65.00% Conservative: 33
Best Local Similarity: 44.38% Mismatches: 52
Query Match: 10.67% Indels: 4
DB: 17 Gaps: 2

US-10-008-355-2 (1-712) x BH388186 (1-537)
Qy 134 LysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeu 153
Db 489 CGTATCACTTCTAAACTGAAACAATGATATGAGC-----GCTGATCAGAGAAACCTATT 436
Qy 154 ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGln 173
Db 435 ATCGATGCTGAAATCAAAGCTATCCAGTCTGAAAACCTCTGAGAAATGGA-----AAATAC 382
Qy 174 LeuCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspVal 193
Db 381 ACTGTAGTGGTGAAGGATTTCTTCAAGGAAATGAGTTTACTACTTTGTATTCAGGAT 322
Qy 194 PheLysAspValArgMetValPheAlaProSerSerValGlyLysPheGlyGlyAsp 213
Db 321 TTCAAAGATGTTCGTCTGGTACTCCGCGTCATCTATTGGTAATATGGTGGTAT 262
Qy 214 ThrAspAsnTyrMetTyrProArgHisThrGlyAspPheSerValPheArgValTyrAla 233
Db 261 ACTGATACTGGGAATGCCAAGACACACAGAGACTTCTCTGTTTCCGTTGTTATGCT 202
Qy 234 GlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyr 253
Db 201 GATAAAATGCAATCTCTGTAATATTCAGCTAACACGTTCCCTTTAAAGCCTAAGCAT 142
Qy 254 PheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhe 273
Db 141 CATTTACCAATTTCTACTAAAGGAAATTAAGCCTGGTATTTGCAATGATTGTAGGATAT 82
Qy 274 ProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGlu 293
Db 81 CTGTGTACAACAACAGATACCTTACTTCTTTCGGAATCGAACAATCGTAAGCAACAT 22

RESULT 4
BH375201 418 bp DNA linear GSS 10-DEC-2001
LOCUS BH375201
DEFINITION AG-ND-173C3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-173C3,
DNA sequence.
ACCESSION BH375201
VERSION BH375201.1 GI:17321343
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 418)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII

```





```

BH400391/c
LOCUS      AG-ND-147H4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4,
DEFINITION DNA sequence.
ACCESSION  BH400391
VERSION    BH400391.1 GI:17346607
KEYWORDS   GSS.
SOURCE     African malaria mosquito.
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.

REFERENCE  1 (bases 1 to 426)
AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL   Unpublished (2001)
COMMENT   Other GSSs: AG-ND-147H4.TR
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjlloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 For
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..426
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-147H4"
                     /clone_lib="ND-TAM"
                     /note="vector: pECBAC1; Site_1: HindIII"

BASE COUNT          111 a 79 c 76 g 160 t
ORIGIN
1
...
Alignment Scores:
Pred. No.:          3.42e-29          Length:          426
Score:              314.00           Matches:          65
Percent Similarity: 65.41%           Conservative: 22
Best Local Similarity: 48.87%        Mismatches:       40
Query Match:        8.44%            Indels:           6
DB:                  17              Gaps:             2

US-10-008-355-2 (1-712) x BH400391 (1-426)

Qy 544 GluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeu 563
Db 399 CAAAAAACACCGTTATTTCATGCGCGCAATTATAAATCTATGCGCTGAATAAAATTC 340
Qy 564 ProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIle----- 578
Db 339 TATCCGGATGCTAAGTCTACGATCAGGTTAACTTATGTCAGATCGAGACGCTTCTCTAAA 280
Qy 579 LysGlyTyrGluProGlnAspGlyAlaThrTyrAsnTyrHisThrGlyLysGlyVal 598
Db 279 AGAGCAGACAGACAGACTACACAGGTATCAGCGAGAATTACTATACCAATCGAAGGTATG 220
Qy 599 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 618
Db 219 ATTAGAGTACACAGAAGGTGACGAGAATTCGATCTTCCACAGGACTCTCTTGATCTT 160
Qy 619 PheArgThrLysAsnTyrGlyArgTyrAlaGlu----AsnGlyGlnLeuHisIleAlaPhe 637

```

```

Db 159 TATAAAAGAAAATTATGTGTATGTACAAAGACAAAGCGCAACTTCATGTAAACTTC 100
Qy 638 LeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGly 657
Db 99 CTTTCTAATAACGATATATACAGGAGGTAATCAGGTCTCCAATTATCGATGGTTACGGA 40
Qy 658 ArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAla 670
Db 39 AGACTTATAGTCTTGCAATTTGACGGAACACAGTGAAGCT 1

RESULT 7
AW774807
LOCUS      AW774807
DEFINITION EST333958 KV3 Medicago truncatula cDNA clone pKV3-24C3, mRNA
sequence.
ACCESSION  AW774807
VERSION    AW774807.1 GI:7718724
KEYWORDS   EST.
SOURCE     barrel medic.
ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE  1 (bases 1 to 398)
AUTHORS   VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L.,
            Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
            Fraser,C.M.
TITLE     ESTs from roots of Medicago truncatula after Rhizobium inoculation
JOURNAL   Unpublished (1999)
COMMENT   Contact: VandenBosch K
            Department of Plant Biology
            University of Minnesota
            220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA.
            Tel: 612 624 2755
            Fax: 612 625 1738
            Email: kvandenb@csb.umn.edu
            Texas A&M EST name:T258226e
            TIGR sequence name:MTEBG14TK
            More information is available at:
            http://chryslie.tamu.edu/medicago
            Seq primer: SKmod (CTA GAA CTA gtg gat CC).
            Location/Qualifiers
     source           1..398
                     /organism="Medicago truncatula"
                     /cultivar="genotype Al7"
                     /db_xref="taxon:3880"
                     /clone="pKV3-24C3"
                     /clone_lib="KV3"
                     /tissue_type="Seedling roots"
                     /dev_stage="3 days post-inoculation with Sinorhizobium
                     meliloti"
                     /lab_host="E. coli strain XL0LR"
                     /note="vector: phluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
                     was directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XL0LR cells."

BASE COUNT          137 a 64 c 81 g 116 t
ORIGIN
1
...
Alignment Scores:
Pred. No.:          1e-19          Length:          398
Score:              239.50         Matches:          48
Percent Similarity: 59.52%         Conservative: 27
Best Local Similarity: 38.10%      Mismatches:       48
Query Match:        6.44%          Indels:           3
DB:                  10           Gaps:             1

US-10-008-355-2 (1-712) x AW774807 (1-398)

```

```

QY 23 LysAlaAspLysGlyMetThrLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMet 42
Db 28 AGAGCCGATGAGGAATGGTCTCTGATGTTATCGAAGAGATTAAATCACAGGATATG 87
QY 43 ArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysProSerIle 62
Db 88 CAAAAAATGGCTTGAATTTACAGCGGAGAGAAATTTACAGCATTAACAATCACAGCTTA 147
QY 63 AlaAsnAlaValIlePheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGly 82
Db 148 AAAGATGCGATTGTACAATTTCAATGGGGTGTCTACTGCGAGAAATCGTTTCAACACGCGT 207
QY 83 LeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAsp 102
Db 208 TTGGTCTTACCAATCACCACTGGTGTATACGGGATTCAGAGACTTTCACACTGCAGAA 267
QY 103 HisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIlePro 122
Db 268 CAAATCATTTAAAAAATGTTTGGGCAAAAAATCGTTTCAGAGAACTAAAAACCAAAA 327
QY 123 GlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGln 142
Db 328 TCTTTA-----TACGTTCGTTTTTTCGTACGTATGCGATGATGTTTCTTAAAGAAAT 378
QY 143 LeuLysGlyIleThrAsp 148
Db 379 TTGTCAAAGTAATGAT 396

```

```

RESULT 8
BH391233
LOCUS AG-ND-141L2.TF ND-TAM Anopheles gambiae genomic clone AG-ND-141L2,
DEFINITION DNA sequence.
ACCESSION BH391233
VERSION BH391233.1 GI:17337374
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.

```

```

REFERENCE 1 (bases 1 to 721)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-141L2.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1..721
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-141L2"
/vector="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

```

## FEATURES

source

```

BASE COUNT 269 a 122 c 147 g 183 t
ORIGIN
Alignment Scores: 1.33e-12 Length: 721
Score: 187.50 Matches: 62
Percent Similarity: 43.10% Conservative: 38
Best Local Similarity: 26.72% Mismatches: 108
Query Match: 5.04% Indels: 24
Gaps: 5
US-10-008-355-2 (1-712) x BH391233 (1-721)
QY 307 IleTyrPLeuAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyr 326
Db 23 GTTATGAAGACATATGATAAGACGATGCTACAGATTAGCTTACGGCTCTAACTAT 82
QY 327 AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeu 346
Db 83 GCAAGTGTTCCTAACTATTGGAAAAACAGAGCCGGAACAAATTTGAAGCGGTTTACAAAAAC 142
QY 347 AspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsn 366
Db 143 GTATCAATTTGGTGCACAAAAAGAGTTGAGAAAAAATACCAACAATGGGCTCATAAAGCT 202
QY 367 GlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGly 386
Db 203 GAAAAACAAGCTGTTTATGGAATGTTATAGCAATACTGATGCATACCAAGCAAAAT 262
QY 387 AlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluVal 406
Db 263 TCTAACAGAAATATCGAAAAAATTTATGTGCACAGTTCCAGAGAAATGCTAAATACATC 322
QY 407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr-----AsnProAspAla 422
Db 323 AGAAATTTCTTCCAGATCGAGATGCTTTAACAGCTACATGAAGCAGATGCATCTGCT 382
QY 423 HisAlaGlyIleLeuLysSerLeuAsp-----AspLysTyrLysAspTyrLeu 438
Db 383 CAGGCAGCTATGAACCTAAATTAGAAGCAGCTGTTAAGCAGGCTTATCAAGGTTTAAT 442
QY 439 ProSerLeuAspArgLysValLeuPro-AlaMetLeuAspIleValArgArgIlePr 458
Db 443 ACTCAACTTGAAGCAGAAATGCTCTCAGAAATGCTTTTATCAATCAAAAAGGTGC 502
QY 458 AlaAspLysLeuProAspIlePheLys-----AsnValIle 470
Db 503 AGCAGATGTTGCTTCTGCAACTGTAATAATCTGTAATGTCATCAGAACTTGCAGAAATATTC 562
QY 470 eAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerVa 490
Db 563 ACAGTCTTCAATTCCTGACAGCCAGCTTCAGTAATTAATCTCCTGACCAATCCAGCGC 622
QY 490 lValProTyrSerAsp-----LysPheHisAlaMetLeuLysSerMetAspLys 506
Db 623 TGAATAATTAGCAATGATAAACTTTTACAAGTTTGTGTCAGGCTATATCGAGACAAACA 682
QY 506 s-----GluLysPheAlaLysAlaIleGlu 514
Db 683 AGTATTTGGCTGAAAAATATGCTAAAAACTGATGAA 716
RESULT 9
BH394089
LOCUS BH394089 706 bp DNA linear GSS 11-DEC-2001
DEFINITION AG-ND-150P20.TF ND-TAM Anopheles gambiae genomic clone AG-ND-150P20
, DNA sequence.
ACCESSION BH394089
VERSION BH394089.1 GI:17340230
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

```

[illegible]

```

Db 427 GCTAAAGCAGCGCTTATGAACAACTTAGCA-----CAGAAAAACCCAGGAGTGA 377
Qy 370 AlavalTyrglyAspValLeuSerLeuGluLysAlaThrArgIleLysTyrAlaSerLysTyr 389
Db 376 AAGCAACTGTTGATGATAAATTAGCCAAATTG-----TATAACGAGCAGGCTCTTTAT 326
Qy 390 AsnArgGluMetThrTyrLeu-----SerGluThrLeuPheGlyGly 403
Db 325 GCATTGAACAATACCTATATCTACTAGGTTACAAAAATGCAGACACATTCGCTCTTGCC 266
Qy 404 ThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHis 423
Db 265 AACTATTTTCATCATATATTTTCAGGAGGTTGAATCCGAAAAAGCTACCCCGGAATCTACT 206
Qy 424 AlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAspArg 443
Db 205 AGAAACTAAAGAACACTTTTAACTTCTTCTACAGGATTATGAAGAGAACTCCGCTGAATTTTACCT 146
Qy 444 LysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAspLysLeuPro 463
Db 145 AAGGTAAACAGCAAAATTTATGCTTTATATGCACAGAAACTCCGCTGAATTTTACCT 86
Qy 464 -----AspIlePheLysAsnVal 469
Db 85 TCAGGATTTGCTCAGTTCAGTGTATGAACAAAAACCTT 47

RESULT 11
BH381877 769 bp DNA linear GSS 10-DEC-2001
LOCUS AG-ND-155J17.TR ND-TAM Anopheles gambiae genomic clone AG-ND-155J17
DEFINITION , DNA sequence.
ACCESSION BH381877
VERSION BH381877.1 GI:17328019
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
REFERENCE Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
AUTHORS Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other GSSs: AG-ND-155J17.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All BAC was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
FEATURES
Location/Qualifiers
source 1..769
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-155J17"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 286 a 124 c 156 g 203 t
ORIGIN

```

```

Alignment Scores: 3.29e-11 Length: 769
Pred. No.: 177.00 Matches: 63
Score: 41.84% Conservative: 37
Percent Similarity: 26.36% Mismatches: 87
Best Local Similarity: 4.76% Indels: 53
Query Match: 17 Gaps: 6
DB:
US-10-008-355-2 (1-712) x BH381877 (1-769)
Qy 307 IleTPrLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyr 326
Db 21 GTTATGAGAGCATATGGATAAAGCAGGATGCTACAAGATTAGCTTACGCGCTCAACTAT 80
Qy 327 AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeu 346
Db 81 GCAAGTGTGTCTTAACCTATTGGAAAAACAGAGCCGCAACAAATTTGAAGCGGTTTACAAAAAC 140
Qy 347 AspValIleGlyArgLysArgAlaGluArgAlaPheAlaAspTrpIleArgLysAsn 366
Db 141 GGAACAATTTGGTGACAAAAAGAGTTGAGAAAAAATTTTATTA-TGGGCTGATAAGCT 199
Qy 367 GlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGly 386
Db 200 GAAAAACAAAGCTGTTTATGCAAAATGATTAGCAAAATCTGATGCATCTACTACAAGCAAAAT 259
Qy 387 AlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluVal 406
Db 260 TCTAACAGAAATATTCGAAAAAATTTATGTCACAGATTCCAGAGAAATGCTAAATACATC 319
Qy 407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr-----AsnProAspAla 422
Db 320 AGAAATTTCTTCCAGATCGGAGATGCCCTTAACAAGCTACATCAACAGCAGGATGCATCTGCT 379
Qy 423 HisAlaGlyIleLeuLysSerLeuAsp-----AspLysTyrLysAspTyrLeu 438
Db 380 CAGGAGCAGCTATGAAACCTAAATTAGAAGCAGCTGTTAAGCAGGCTTATGAAGGTTTAAAT 439
Qy 439 ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIlePro 458
Db 440 ACTCAACTTGAAGCAGAAATGCTATCTCAGATGGCTTTTATATATCAATCAAAAAGTAGCA 499
Qy 459 AlaAsp-----LysLeuProAspIlePhe 466
Db 500 GCAGATGTTGCTTCTCCAACTGTAAATCTGTAANTGCATCAGAACTTCGGAATATTGCA 559
Qy 467 LysAsnValIle----- 470
Db 560 CAGTCTTCAATTTTGGCAAAACGAGCTTCAGTAATTAACCTCCTGAAACAATCCAAGCGCT 619
Qy 471 -----AspLysLysPheLys-----GlyAspThrLys 479
Db 620 GAAAAATTAGCGAATGATAAACTTTTACAAGTTTGTGCGAGGTATATCGGAGACACAAG 679
Qy 480 LysTyr-----AlaAspPheValPheAspLysSerValValProTyr 493
Db 680 TATTGCTCAAAATATGCTAAACTGATGAAGGTTTCAGAAAACACAGCGCTTTATAT 736

```

```

RESULT 12
BH164253/1
LOCUS ENT7G63TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION BH164253
VERSION BH164253.1 GI:15737691
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE Eukaryota; Entamoebidae; Entamoeba.
AUTHORS 1 (bases 1 to 1005)
TITLE Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica

```

JOURNAL COMMENT

HMI-IMSS sheared DNA library (2001)  
Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 15  
High quality sequence stop: 487.

FEATURES  
source  
1..1005  
Location/Qualifiers  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOSt; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)." 169 t

BASE COUNT  
186 a 314 c 336 g 169 t

ORIGIN

Alignment Scores:  
Pred. No.: Length: 1005  
Score: 160.50 Matches: 52  
Percent Similarity: 53.0% Conservative: 18  
Best Local Similarity: 39.3% Mismatches: 53  
Query Match: 4.32% Indels: 9  
DB: 17 Gaps: 5

US-10-008-355-2 (1-712) x BH164253 (1-1005)

Qy 10 LeuGlyAlaAlaLeuLeuGly-----AlaSerGlyValAlaLysAlaAspLys 26  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 393 CTTCACCGCCGTGGCCCTGGCCCTGCAGCCTCGCCGCCGT---GCCACCGCATGA 337  
  
Qy 27 GlyMetTripleuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGly 46  
||||||| : : : : : ||| : : : : : ||| : : : : : |||  
Db 336 GCATGTGGATGCCAGCAGTG---CCGAGCTGCCAACCGCTGCAGCCCGCAGGT 280  
  
Qy 47 PheThrLeuProLeuAspSerLeuTySerPheAspLysProSerIleAlaAsnAlaVal 66  
||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||  
Db 279 TTCAAGGGCAATCCGCGCACCTGCCAATGTACCGCGCCGCGGT---AGCGCCGTG 223  
  
Qy 67 ValIlePheGlyGlycysThrGlyIleThrValSerAspGlnGlyLeulePheThr 86  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 222 GTCCGCCCGTGGTGGT---ACC GGTTGGTTCGTCTGTCGCCGCGCGCTGCTGACC 166  
  
Qy 87 AsnHisGlycysGlyTyrglyAlaIleGlnSerGlnSerThrValAspHisAspTyrl 106  
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 165 AACACACACGTGGCGTGGTGTACAGTAGTACACAGCTCCGCCGCGCACCTGATPC 106  
  
Qy 107 ArgAspGlyPheValSerArgThrMetGlyGluGluLeuProfile--ProGlyLeuSerV 126  
||||||| : : : : : ||| : : : ||| : : : ||| : : : ||| : : : |||  
Db 105 ACTGGTGCTTCATCGCCAAACGCCCTGCCGAGCGCCGCCCAACCGGACTTCGCC 46  
  
Qy 126 alLysTyrlLeuArgLysIleValLysValThr 136  
||||||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 203 GAAACAAAGCTGGTTATGAGAAATCTATTAGCAAACTAGTGCATACATAAGCAA 260
RESULT 14
BH371846
LOCUS
DEFINITION AG-ND-162M17.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162M17
, DNA sequence.
ACCESSION BH371846
VERSION BH371846.1 GI:17317971
KEYWORDS GSS:
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE 1 (bases 1 to 822)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other GSSs: AG-ND-162M17.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
FEATURES
Source
1..822
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-162M17"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 235 a 142 c 155 g 290 t
ORIGIN
Alignment Scores:
Pred. No.: 6,46e-06 Length: 822
Score: 136.00 Matches: 27
Percent Similarity: 79.07% Conservative: 7
Best Local Similarity: 62.79% Mismatches: 9
Query Match: 3.66% Indels: 0
DB: 17 Gaps: 0
US-10-008-355-2 (1-712) x BH371846 (1-822)
Qy 670 AlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIle 689
|||||
Db 2 GCTTTAAGCGGTGACATGTTTGTGAACCTAAATACAAAGAACGATTAAACGTAGACGTT 61
|||||
Qy 690 ArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeuIleGlnGluLeu 709
|||||
Db 62 AGATACGTACTTTGGGTATCGACAAAGTTTGCAGGTGCTAAAAAAGCTTAATAGCGAATTG 121
|||||
Qy 710 LysLeuIle 712
Db 122 ACTTTAGTA 130
|||||
RESULT 15
AY107109

```

```

LOCUS AY107109
DEFINITION Zea mays PC0121171 mRNA sequence.
ACCESSION AY107109
VERSION AY107109.1 GI:21210187
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1869)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1869)
AUTHORS Coe,E.C.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
FEATURES
source
1..1869
Location/Qualifiers
/organism="Zea mays"
/db_xref="MaizeDB:637050"
/db_xref="taxon:4577"
/clone="PC0121171"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACS in conjunction with the Maize
Mapping Project"
BASE COUNT 547 a 337 c 473 g 512 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0092 Length: 1869
Score: 116.50 Matches: 105
Percent Similarity: 34.17% Conservative: 73
Best Local Similarity: 20.15% Mismatches: 180
Query Match: 3.13% Indels: 163
DB: 11 Gaps: 27
US-10-008-355-2 (1-712) x AY107109 (1-1869)
Qy 82 GlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThrVal 101
|||||
Db 234 GCCCCTTTGGCATCATCACACTTCACACAGAGGGGTAGAGTATTTAGTCAATTCCTT 293
|||||
Qy 102 AspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIle 121
|||||
Db 294 -----TTTGTCTCTGCTACAAAGGAGGAGATGTAACCTGCT 329
|||||
Qy 122 ProGlyLeuSerValLysTyrLeuArgLysIleValLys-----ValThrAspLys 138
|||||
Db 330 GATAGTAGAAGAACCAAGAAATATCAGACTATATGTTAAGCGGTGATTTCATATCAGATGAC 389
|||||
Qy 139 ValGluGlyGlnLeu-----LysGlyIleThrAspGluMet 150
|||||
Db 390 TTTGATGGAGAAATCTTCCCAAGATATTTGAGCTTTTGCAAGGGTGTCTTCTCAGCTCAAT 449
|||||
Qy 151 Glu-----ArgLeuArgLysAlaGlnGluValCysGlnGlu----- 162
|||||
Db 450 GACCTTCATTTGAATGTTTCTCGTGAGATTTCTTCAAGAAAGTCGAATTTCTCGTATAATG 509
|||||
Qy 163 -----LeuAlaLysLys-----GluAsn 168
|||||
Db -510 AGAAAAAGCGTCGTGCGAAAGACCCCTTGTATGATGATGATGATGATGATGATGATGATGAT 569
|||||
Qy 169 AlaAspGluAsnGlnGlnCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeu 188

```

Db	1413	A	TCCACGACGACGAC-----GCACCTGTGAGGTCA---GAGCAGGCCCGCTTTGAGCT	1460
Qy	511	LysAlaIleGluLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAArg	530	
	:	:: :::::	: : :	
Db	1461	CAGTGGTGAGGACGACAGCCC---GTTGAGGCTGGTCAGCAGAGTGAAGGCCCTGTTCTG	1517	
Qy	531	Ala	531	
	:	:: :::::	: : :	
Db	1518	GCA	1520	

Search completed: May 23, 2003, 12:11:22  
Job time : 2030.73 secs





GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:41:33 ; Search time 38.2125 Seconds  
(without alignments)  
548.227 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MQWKLKILGALLGASG.....LFMDKMGQCPRIQLKLI 712

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	3.5	883	2	US-08-953-492-2
2	127.5	3.4	1194	2	US-08-488-940-18
3	123.5	3.3	713	4	US-08-039-584-53
4	122.5	3.3	800	2	US-08-488-940-4
5	122.5	3.3	813	2	US-08-488-940-3
6	122.5	3.3	1181	2	US-08-488-940-2
7	122.5	3.3	1194	2	US-08-488-940-1
8	122.5	3.3	1194	2	US-08-488-940-17
9	110	3.0	1027	4	US-08-446-137B-2
10	109	2.9	1577	2	US-08-793-824-2
11	106	2.9	972	3	US-08-335-844A-23
12	103.5	2.8	1114	4	US-08-811-583-2
13	102.5	2.8	833	4	US-09-514-302-3
14	102.5	2.8	856	4	US-09-134-001C-5438
15	102.5	2.8	888	4	US-09-134-001C-3032
16	102.5	2.8	1938	4	US-09-514-302-2
17	102	2.7	413	4	US-09-134-001C-5111
18	101.5	2.7	747	1	US-07-854-596B-40
19	101	2.7	433	1	US-08-417-492-2
20	101	2.7	772	4	US-09-134-078-28
21	100.5	2.7	638	1	US-08-712-241-6
22	100.5	2.7	885	1	US-08-484-105-14
23	100.5	2.7	885	1	US-08-484-106-14
24	100	2.7	4536	4	US-09-180-422B-27
25	100	2.7	15281	2	US-08-471-119A-2
26	99.5	2.7	722	4	US-08-961-083-84
27	99.5	2.7	849	4	US-09-157-257-4

Sequence 1, Appli  
Sequence 2, Appli  
Sequence 5, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 10, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 10, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 12, Appli

ALIGNMENTS

RESULT 1  
US-08-953-492-2  
; Sequence 2, Application US/08953492  
; Patent No. 5849555  
; GENERAL INFORMATION:  
; APPLICANT: Brown, James  
; APPLICANT: Jaworski, Deborah  
; APPLICANT: Lawlor, Elizabeth  
; APPLICANT: Wang, Min  
; TITLE OF INVENTION: NOVEL vals  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Smithline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/953,492  
; FILING DATE: 17-OCT-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/844,064  
; FILING DATE: 18-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9607991.8  
; FILING DATE: 18-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmil, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P31458-4/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 883 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-953-492-2





```
DB 343 -----NHDDTNRIITYVMG--KRP-----EGENASYHLAY-----D 371
QY 277 TDY-----LTSWGVEDRTENENNPR-----EVGKIQGIWKKEAMSADQAT 318
DB 372 KORYTEEREVYSYLYRTGTPDPNDPNKNSQLVSVAGTVEGTNQDISLKFFEDILTS 431
QY 319 RIKYASKYQ-----SANYWKNSIGNRGLARLDVIGRKAERAFADWIRKN-----366
DB 432 RPAHGKTEQGLSPKSKPEATDSGAMSHKLEKADLL--KAIOEQLIAN--VHSNDDYFEVI 488
QY 367 --GKSNAVYGDVLSLEKAYKEGAKANREMTYLSFTLPGTVEVVRFAQFANALATNPDAHA 424
DB 489 DFASDATITDRNGKVYFADKGS-VTLPTQPVQOEFTYVKNREQAYRINKKSGLEENINN 540
QY 425 GILKSLDDKY-----KDYLPSEL-DRKVLPAMLDIVRRIPADKLPDIFKNVIDKK 473
DB 541 A--KSVDEVYTVQFTPLNPDDEFRLGKDTKLLKTL--AIGDTITSQELLAQAQSLNKN 596
QY 474 FKGDTKKYADFVFDKSVVPYSDFHAMLKSMDEKFAKAEKDPVELSKSVIAAARATQ 533
DB 597 HPG-----YTIYERDSSIVTHNDIFRTILPMDQOEFTYVKNREQAYRINKKSGLEENINN 652
QY 534 ADAMANA-YAIEKKG-----RLFFAGLREMYPGRALPDSANFT---MRMSYSGI 578
DB 653 TDLISEKYVVLKKGKPYDPDRSHLKLFTIKYVDVDTNELLKSEQLLTASERNLDFRDL 712
QY 579 KGYEPDGA--WYN-----YHTTGKGVLEKQDPKSEFAVOENILDLFR--TKNYG 625
DB 713 --YDPRDKAKLLYNLDARFIMDYTLTGK-----VEDNHDDTNRIITYVMG 756
QY 626 R--YAENGOLHIAF 637
DB 757 KRPEGENASYHLAY 770
```

## RESULT 5

```
US-08-488-940-3
; Sequence 3, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 813 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
US-08-488-940-3
Query Match 3.3%; Score 122.5; DB 2; Length 813;
Best Local Similarity 19.3%; Pred. No. 0.022;
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;
QY 75 GITVSDQGL-----IFTNHHGCGYGAIQSQSTVDHDLRDLRGFVSRMTGELPIP---GLS 125
DB 190 GDTITSQELLAQAQSLNKNHFGYTIYERDSSI---VTHNDIFRTI---LPMQOEFTYR 243
QY 126 VKYLRIKIVTKVQEGQLKGIITDEMERLKAQEVCOELAKE-----167
DB 244 VKNREQAYRINKK-----SGLNEEINNTDLISEKYVVLKKGKPYDPDRSHLKLFTIKY 298
QY 168 -NADENQLCIVPEFYSNNEY---FLIVYDVFKDVRMVFAPPSSVGKFG-----GDTON 216
DB 299 VVDVTNELLKSEQLLTASERNLDFRDLVDPDKAKLLY---NNLDAFGIMDYTLTGKVED 355
QY 217 WMPRHITGDFSVERVYAGADNRPAEYSKDKPKYPVYFAAVSMQGYKADDYAMTIGFPGS 276
DB 356 ----NHDDTNRIITYVMG--KRP-----EGENASYHLAY-----D 384
QY 277 TDY-----LTSWGVEDRTENENNPR-----EVGKIQGIWKKEAMSADQAT 318
DB 385 KORYTEEREVYSYLYRTGTPDPNDPNKNSQLVSVAGTVEGTNQDISLKFFEDILTS 444
QY 319 RIKYASKYQ-----SANYWKNSIGNRGLARLDVIGRKAERAFADWIRKN-----366
DB 445 RPAHGKTEQGLSPKSKPEATDSGAMSHKLEKADLL--KAIOEQLIAN--VHSNDDYFEVI 501
QY 367 --GKSNAVYGDVLSLEKAYKEGAKANREMTYLSFTLPGTVEVVRFAQFANALATNPDAHA 424
DB 502 DFASDATITDRNGKVYFADKGS-VTLPTQPVQOEFTYVKNREQAYRINKKSGLEENINN 553
QY 425 GILKSLDDKY-----KDYLPSEL-DRKVLPAMLDIVRRIPADKLPDIFKNVIDKK 473
DB 554 A--KSVDEVYTVQFTPLNPDDEFRLGKDTKLLKTL--AIGDTITSQELLAQAQSLNKN 609
QY 474 FKGDTKKYADFVFDKSVVPYSDFHAMLKSMDEKFAKAEKDPVELSKSVIAAARATQ 533
DB 610 HPG-----YTIYERDSSIVTHNDIFRTILPMDQOEFTYVKNREQAYRINKKSGLEENINN 665
QY 534 ADAMANA-YAIEKKG-----RLFFAGLREMYPGRALPDSANFT---MRMSYSGI 578
DB 666 TDLISEKYVVLKKGKPYDPDRSHLKLFTIKYVDVDTNELLKSEQLLTASERNLDFRDL 725
QY 579 KGYEPDGA--WYN-----YHTTGKGVLEKQDPKSEFAVOENILDLFR--TKNYG 625
DB 726 --YDPRDKAKLLYNLDARFIMDYTLTGK-----VEDNHDDTNRIITYVMG 769
QY 626 R--YAENGOLHIAF 637
DB 770 KRPEGENASYHLAY 783
RESULT 6
US-08-488-940-2
; Sequence 2, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,940  
FILING DATE: 09-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/009001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1181 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-488-940-2

Query Match 3.3%; Score 122.5; DB 2; Length 1181;  
Best Local Similarity 19.3%; Pred. No. 0.04;  
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

```
QY 75 GITVSDQGL-----IFTNHHCYGAISQSTVDHYLDGFSVRTMGEEPLIP---GLS 125
Db 558 GDTITSQELLAAQAQSIILNKNHPGYTIYERDSSI---VTHDNDIFRTI---LPMDOEFTYR 611
QY 126 VKYLRKIVKVTDKVGEQLKGTIDEMERLKAQEVCOELAKKE----- 167
Db 612 VKNRQAYRINKK-----SGLNEEINNTDLISEKYVVLKKGKPYDPDRSHLKLFTIKY 666
QY 168 -NADENQICIVEPFYSNNEY---FLIVVDVFKDVRMVFAPPSSVGKFG-----GDTDN 216
Db 667 VDVDTNELLKSEQLLTASERNLDFRDLDPDRKAKLLY---NNLDAFGIMDYTLTGKVED 723
QY 217 WMPRHTGDFSFRVYAGADNRPAYSKDNKPKVPYFAAASMOQYKADDYAMTIGFPGS 276
Db 724 ----NHDDTNRIITYMG--KRP-----EGENASYHLAY-----D 752
QY 277 TDRY-----LTSWGVEDRIENENPRI-----EVRGIKQGIWKEMSADQAT 318
Db 753 KDRYTEEREVYSYLRYGTPTIPDNPKNNSQLVSVAGTVEGTNQDISLKFFEDILTS 812
QY 319 RIKYASKYAO-----SANYWKNISGMNRLARLDVIGRKRAEERAFADWIRKN----- 366
Db 813 RPAHGGKTEQGLSPKSPKPFATDSGAMSHKLEKADLL--KAIOEQLIAN-VHSNDDYFEVI 869
QY 367 --GSAVYGDVLSLEKAYKEGAKARNTYLSFTLFGTEVVRFAQFANALATNPDAHA 424
Db 870 DFASDATITDRNGKYVFADKDG--VTLPTQPOVEELLGSHVVRVYKE-----KPIQNO 921
QY 425 GILKSLDKY-----KDYLP--DRKVLPAMLDIVRRIRIPADKLPIFKNVIDKK 473
Db 922 A--KSDVDEYTVQFTPLNPDODDFRGLKDTLLKTL--AIGDTITSQELLAAQAQSIILNKN 977
QY 474 FKGDYKTKYADVFVDSKVPYSDKFKHMLKSMDEKFAKAEKADPAVELSKSVIAAARIQ 533
Db 978 HPG-----YTIYERDSSIVTHNDIFRTILPMDQEFYTVKNEQAYRINKSGLNEEINN 1033
QY 534 ADAMANA-YATEKKG-----RLFFAGLRMYGCRALPDSANET---MRMSYGS 578
Db 1034 TDLISEKYVVLKKGKPYDPDRSHLKLFTIKYVDVDTNELLKSEQLLTASERNLDFRDL 1093
QY 579 KGYPEQDGA--WYN-----YHTTGKGVLEKQDPKSDFAVOENILDLFR---TKNKG 625
Db 1094 --YDRPKAKLLYNLNDLAFGIMDYTLTGK-----VEDNHDDTNRIITYMG 1137
QY 626 R--YAENGQLHAF 637
```

Db 1138 KRPEGENASYHLAY 1151  
RESULT 7:  
US-08-488-940-1  
Sequence 1, Application US/08488940  
Patent No. 5854049  
GENERAL INFORMATION:  
APPLICANT: Reed, Guy L.  
TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,940  
FILING DATE: 09-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/009001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-488-940-1

Query Match 3.3%; Score 122.5; DB 2; Length 1194;  
Best Local Similarity 19.3%; Pred. No. 0.04;  
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

```
QY 75 GITVSDQGL-----IFTNHHCYGAISQSTVDHYLDGFSVRTMGEEPLIP---GLS 125
Db 571 GDTITSQELLAAQAQSIILNKNHPGYTIYERDSSI---VTHDNDIFRTI---LPMDOEFTYR 624
QY 126 VKYLRKIVKVTDKVGEQLKGTIDEMERLKAQEVCOELAKKE----- 167
Db 625 VKNRQAYRINKK-----SGLNEEINNTDLISEKYVVLKKGKPYDPDRSHLKLFTIKY 679
QY 168 -NADENQICIVEPFYSNNEY---FLIVVDVFKDVRMVFAPPSSVGKFG-----GDTDN 216
Db 680 VDVDTNELLKSEQLLTASERNLDFRDLDPDRKAKLLY---NNLDAFGIMDYTLTGKVED 736
QY 217 WMPRHTGDFSFRVYAGADNRPAYSKDNKPKVPYFAAASMOQYKADDYAMTIGFPGS 276
Db 737 ----NHDDTNRIITYMG--KRP-----EGENASYHLAY-----D 765
QY 277 TDRY-----LTSWGVEDRIENENPRI-----EVRGIKQGIWKEMSADQAT 318
Db 766 KDRYTEEREVYSYLRYGTPTIPDNPKNNSQLVSVAGTVEGTNQDISLKFFEDILTS 825
QY 319 RIKYASKYAO-----SANYWKNISGMNRLARLDVIGRKRAEERAFADWIRKN----- 366
Db 826 RPAHGGKTEQGLSPKSPKPFATDSGAMSHKLEKADLL--KAIOEQLIAN-VHSNDDYFEVI 882
```

```
QY 367 --GKSAYGVGVVSSLEKAYKEGAKANRENTYLSLFGTGVVVRFAQFANALATNPDAHA 424
Db 883 DFASDATITDRNGKVYFADKGS-VTLPTQPVQVEFLLSGHVRVRYKE-----KPIQNO 934
QY 425 GILKSLDDKY-----KDYLSL-DRKVLPAMLDIVRRRIPADKLPDIFKNVIDKK 473
Db 935 A--KSDVVEYVQFTPLNPDDEFRLGKDKTLKLT--AIGDTITSQELLAAQASILNKN 990
QY 474 FKGDTKKYADFVDFKSVVPSYDKFHAMLSMDKFAKAEKDPAVELSKSVIAAARATQ 533
Db 991 HPG-----YTIYERDSSIVTHDNDIFRTILPMQDEFTYRVKNREQAYRINKKSGLNBEINN 1046
QY 534 ADAMANA-YAIEKKG-----RLFFAGLREMPGRALPSDANFT---MRMSYSGI 578
Db 1047 TDLISEKYYVVKKGKPYDPDRSHLKLFTIKYVDVDTNELKSEQLLTASERNLDFRDL 1106
QY 579 KGYPEODGA--WYN-----YHTTGKGVLEKQDPKSDFAVOENLIDLFR--TKNYG 625
Db 1107 --YDPRDKAKLLYNLDAFGIMDYTLTGK-----VEDNHDDTNRIITVYMG 1150
QY 626 R--YAENGOLHIAF 637
Db 1151 KRPEGENASYHLAY 1164

RESULT 8
US-08-488-940-17
; Sequence 17, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1194 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-940-17

Query Match 3.3%; Score 122.5; DB 2; Length 1194;
Best Local Similarity 19.3%; Pred. No. 0.041;
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

QY 75 GITVSDQGL-----IFTHNHGCGYGAIQSQSTVDHDLRDLRGVSRMTGMEELP---GLS 125
Db 571 GDTITSQELLAAQASILNKNHPGYTIYERDSSI---VTHDNDIFRTI---LPMQDEFTYR 624
```

```
QY 126 VKYLKRVKVTDKVEGOLKGIITDEMERLKRKAQEQELAKKE----- 167
Db 625 VKNRQAYRINK-----SGLNEEINNTDLISEKYVVLKKGKPYDPDRSHLKLFTIKY 679
QY 168 -NADENQICLIEPFYSNNEY---FLIVVDVKDVRMVFAPSSVKGKFG-----GDTDN 216
Db 680 VVDVTNELKSEQLLTASERNLDFRDLDPDRKAKLLY---NNLDAFGIMDYTLTGKVED 736
QY 217 WMPRHITGDFSVFRYAGADNRPAEYSKDNKPKVPYFAAVSMQGYKADDYANTTIGFPGS 276
Db 737 ----NHDDTNRIITVYMG---KRP-----EGENASYHLAY-----D 765
QY 277 TORY-----LTSWGVEDRIENENPRI-----EVRGIKQGIWKKEAMSADOAT 318
Db 766 KDRYTEEREVEVYSIRYGTPTIPDNPNKNSQLVSVAGTVEGTNQDISLKEFEIDLTS 825
QY 319 RIKYASKYAO-----SANYKNKSIGMNRGLARLDVIGRKRAERAPADWIRKN----- 366
Db 826 RPAHGKGTEQGLSPKSKPEATDSGAMSHKLEKADLL--KAIQEQLIAN-VHSNDDYFEVI 882
QY 367 --GKSAYGVGVVSSLEKAYKEGAKANRENTYLSLFGTGVVVRFAQFANALATNPDAHA 424
Db 883 DFASDATITDRNGKVYFADKGS-VTLPTQPVQVEFLLSGHVRVRYKE-----KPIQNO 934
QY 425 GILKSLDDKY-----KDYLSL-DRKVLPAMLDIVRRRIPADKLPDIFKNVIDKK 473
Db 935 A--KSDVVEYVQFTPLNPDDEFRLGKDKTLKLT--AIGDTITSQELLAAQASILNKN 990
QY 474 FKGDTKKYADFVDFKSVVPSYDKFHAMLSMDKFAKAEKDPAVELSKSVIAAARATQ 533
Db 991 HPG-----YTIYERDSSIVTHDNDIFRTILPMQDEFTYRVKNREQAYRINKKSGLNBEINN 1046
QY 534 ADAMANA-YAIEKKG-----RLFFAGLREMPGRALPSDANFT---MRMSYSGI 578
Db 1047 TDLISEKYYVVKKGKPYDPDRSHLKLFTIKYVDVDTNELKSEQLLTASERNLDFRDL 1106
QY 579 KGYPEODGA--WYN-----YHTTGKGVLEKQDPKSDFAVOENLIDLFR--TKNYG 625
Db 1107 --YDPRDKAKLLYNLDAFGIMDYTLTGK-----VEDNHDDTNRIITVYMG 1150
QY 626 R--YAENGOLHIAF 637
Db 1151 KRPEGENASYHLAY 1164

RESULT 9
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B
```

```
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McWaters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-446-137B-2

Query Match          3.0%; Score 110; DB 4; Length 1027;
Best Local Similarity 20.6%; Pred. No. 0.47;
Matches 118; Conservative 63; Mismatches 206; Indels 186; Gaps 29;

Qy 29 WLLNELNOENLDRMELGFTLPLDLSYFDPKPSIANAVVIFGGCTGITVSDOGLIFTNH 88
Db 587 WLLNKAEEAEKELKEAGIT---SDLY-FSLINKAKTV-----EGVEALKNEIL--KA 633
Qy 89 HCGYGAIQSOSTVDHYLDYRDGFSRTMGEELPIPLSVKYLKIVKTVKVEGO---LK 144
Db 634 HAG-----EETPE---LKDCYATVEEAAEAAKALKNDVNNAYEIVQGDGRYYVLK 684
Qy 145 -GITDEMERLKAQEVQC-----ELAKKENADENQLCIVEFFYSNNEFLI----- 189
Db 685 IEVADDEEPGDTPEVQGYATYEEAAEAAKALKEDKV-----NNAYEVVQGDGR 736
Qy 190 VYDVEK-DVRMVFAPPSSVSGKFGDGT-DNWMVPRHTGDSVFRVYAGADNRPAYSKDNK 247
Db 737 YYYVLKIEDKEDEQPGEPGENPGITIDWLLKNKED-----ALKELKEAGISSD-- 787
Qy 248 PYKPYFAVSMQYKADDYAMTIGFPSTDRYLTSMGVEDRIENE-----NNPR 297
Db 788 ---IYFDAI-----NKAKTVEGVE-ALKNEILKAHAERKPGENPG 822
Qy 298 IEV-----RGIKQGIWE-----AMSADATRIKYASKYAQSANYWKNSI-----GMN 340
Db 823 ITIDWLLKNKAEAAIKELKEAGITAEYFLNLINKAKTVEGESLKNELKAHAERKGEN 882
Qy 341 RGLARLDVIGRKAEERAFADWIRKNGKSAVYGDVLSLEKAYKEGAKANREMYTSLTL 400
Db 883 PGI-----TIDWLLKNK-----EDAIKE-----LKEAGITSIIY 913
Qy 401 FGGTEVVRFAFANAL-----ATNPDAHAGILKSLDDKYDYLPSLDRKYLPAMLDIVR 454
Db 914 FDAINKAKTIEGVEALKNEILKAHKHDEPKGKPEDKPKEDKPKEDK-----963
Qy 455 RRIADKLPLDIFKNVDKFKGDTKYYADFVDSKVPYPSDKFHMLKSMDEKFEAKAIE 514
Db 964 ---PEDKKPGEDKPKEDKK-PKRTDK-----DSPNKKKKAKL-- 996
Qy 515 KDPAVELSKSVIAAARAIAQADAMANAYAIKKGK 547
Db 997 --PRAGSEAEILTAAALSTAAGAYVSLKKPK 1027

RESULT 10
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of plants to
; INCREASE STORED CARBOHYDRATES
; NUMBER OF SEQUENCES: 2
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793.824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
; US-08-793-824-2

Query Match          2.9%; Score 109; DB 2; Length 1577;
Best Local Similarity 20.7%; Pred. No. 1.2;
Matches 145; Conservative 95; Mismatches 262; Indels 198; Gaps 39;

Qy 46 GFTPLDLSL-YSFQKPSIANAVVIFGGCTGITVSDOGLIFTNHHCYGA-IQOSTVDH 103
Db 690 GFTTMDLKAQFE---LYNA-----DIADKKKYQYNIPIAAYATMTLNKDSITR 737
Qy 104 DYLRDGFV--SRTMGEELPIPLSVKYLKIVKTVKVEGOLKITDEMERLKAQEVQC 161
Db 738 VYVGLDFTDDGQYMAEKSPYNAIDALLRARIK-----YVAGGDMKVTKLNGYEIMS 790
Qy 162 ELAKKENADE-NQLCIVEFFYSNNEYFLIVYDVFKDVRMVFAPPSSVSGKFGDGTDNWMP 220
Db 791 SVRYGKGAEEANQLGTAE---TRNQGMLVLTANRPMKML-----GANDRLV-- 833
Qy 221 RHTGDSVFRVYAGADNRPAYSKDNKPKYKPYFAA-----VSMQ 260
Db 834 -----VNMGAHK-----NQAYRPLLKSLKGLATYLCDSVPAGLVRYTDNQ 876
Qy 261 G---YKADYA--WTIGFGPSTDRYLTSMGVEDRIENENNPRIEVGKIQG--IWKEAMS 313
Db 877 GNLFTTADDIAGHSVEVSG---YLAVMVPVGASEQ--DARTKASSTKKGQVPESSAA 931
Qy 314 ADQAT-----RIKYASKY-----AQSANVYKNSIGMNRGLARLDVIGR-KRAER 357
Db 932 LDSQVIYEGFSNFQDFVKTPTSQYTNRVIAQNAKLEK-----EWGITSFEFAPQYVSSQDG 986
Qy 358 AFADWIRKNG-----KSAVYG---DVLSSLEKAYKEGAKANREMYTSLSETLFG 402
Db 987 TFLDSIENGAFEDRYDIAMSKNNKYGLSKDMDALRALHAEGISAIAD--WVDPQIYN 1044
Qy 403 --GTEVVRFAFANALATNPDAHAGILKSL-DDKYDYLPSLDRKVLPAMLDIVRRRIPA 459
Db 1045 LPKGEVVT-ASRTNSYGT-PRPNAETIYSLYAAKTRTEGNDFOGKYGGAFIDELKAKYPA 1102
Qy 460 DKLPDIFKNV-IDKKFGDTKYYADFVDSKSVVYPSDKFHMLKSMDEKFEAKAIEKQDA 518
Db 1103 -----IFERVQISNGRKLTT-----NEKITOWSAKY----- 1128
```

```

Query Match      2.9%; Score 106; DB 3; Length 972;
Best Local Similarity 19.9%; Pred. No. 1;
Matches 103; Conservative 71; Mismatches 223; Indels 120; Gaps 21;

89 HCGYGAIQSQSTVDHDIYLRDGFYSRTMGELPPLGLSVKYLRIKIVKVTDKVEGOLKGITD 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
383 HQWFEGNLVTLKKWDDTWTNLNGEFA--TFVEYLGIMDEIHNHNER--TQDFFLLDGMDRGMRA 438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12  
 US-08-811-583-2  
 : Sequence 2, Application US/08811583  
 : Patent No. 6218142  
 : GENERAL INFORMATION:  
 : APPLICANT: Wasseneberger, Michael  
 : APPLICANT: Riedel, Leonhard  
 : APPLICANT: Schielebel, Winfried  
 : APPLICANT: Sanger, Helmut  
 : TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
 : TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN  
 : NUMBER OF SEQUENCES: 13  
 : NUMBER OF SEQUENCES: 13  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: FISH & NEAVE  
 : STREET: 1251 Avenue of the Americas  
 : CITY: New York  
 : STATE: New York  
 : COUNTRY: USA  
 : ZIP: 10020  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/811,583  
 : FILING DATE: 05-MAR-1997  
 : CLASSIFICATION: 536  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Haley, James F.  
 : REGISTRATION NUMBER: 27,794  
 : REFERENCE/DOCKET NUMBER: MPG-1  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 212-596-9000  
 : TELEFAX: 212-596-9090  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1114 amino acids



; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-811-583-2

Query Match 2.8%; Score 103.5; DB 4; Length 1114;  
Best Local Similarity 19.8%; Pred. No. 2.2;  
Matches 127; Conservative 78; Mismatches 229; Indels 207; Gaps 30;  
Qy 92 YCAIQSQSTVDHDLRDCGFVSRKTGE-----ELPIPLGSVKY-----LRKIVK 134  
Db 556 YGGYGVGVDPDSSMKLSLRSKSKYESDNKLDVLGWS-KYQPCYNLRLLITLLSLG 614  
Qy 135 VTDKV-EQQLGIIDEMERL-----RKAQEVCOELAKKENADENOLCI-----VEPPYS 182  
Db 615 VKDEVLEQKEAEVDQLDAILHDSLKAQEALELMSPGENTWILKMLNCGYKPDAPPLS 674  
Qy 183 NNEFLIVYDFK-----DVR---MVFPAPSSV-----GK 209  
Db 675 -----MMLOTFRASKLLDLRLRSRIFIPNGRTWMGCLDESRTLEYGVFVQFTGAGH 728  
Qy 210 FGGDTNNMWRHRTGDFSVFVYAGADNRPAEYSKDNKPKYKPVYFAAVSMOGYKADYAM 269  
Db 729 FSDDLHPNNSRSINSFILKGNVNVAKNCPCLHPCDIRLKAIVNRALH---HMVD----- 781  
Qy 270 TIGFP-----GST---DRYLTSGVEDRIENENNPRIEVRGIIKOGIWEAMS- 313  
Db 782 CVFPPQKGRPHNPNECSGSLDGDYFVCM-----DQDMIPPRQVQPMEXPPAPSI 832  
Qy 314 -ADQATRIKYASKAQSANYKN-SIGNNRGLARLDVIGRKRAERAPADWIRKNG----- 367  
Db 833 QLDHVDTEEEVEEY--FTNYIVNDSLGI---IANAHVVFADREPDMAMSDPCKLAEFLS 887  
Qy 368 -----KSAYGDLVSSLE-KAYKEGAKANREMYVLSLTFGGTEVVVFAQANALATNP 420  
Db 888 IAVDFPKTGVAEIPSQLRPEYDFDMDKPKDTSYISERVIG----- 929  
Qy 421 DAHAGILKSLDDKYDLPSLDRKVLPAMLDIVRRRIPADKLPDIFKNVIDKFKGDTKK 480  
Db 930 -----KLFKRVKDKAPQAS-SIATFTROVARRSYDADMEVDGFDYIDEAF--DYKT 978  
Qy 481 YADFVDSVVPYSDKPHAML-----KSMDEKFAKAEIKDPDAVELSKSVIAARA 531  
Db 979 EYDNKLGMLMDYIGIKTEAELISGIMKASKTFDRRKDAEALS-----VAVRA 1026  
Qy 532 IQADAMANAIAIEKGRLLFFAGLREMPGRALPSDANFTMRMSYGSIKGYEPDQGAWYN- 590  
Db 1027 LRKEAR-----ANFKRR-----NDIDMLPKASANTHV 1054  
Qy 591 -YHTGKGV-----LEKQDPKSDFAVOENILDLFRTKNYGR 626  
Db 1055 TYHTPYMGYCNQGLKRAHFISFPWCYVDQLIQIKKARNR 1095

RESULT 13  
US-09-514-302-3  
; Sequence 3, Application US/09514302  
; Patent No. 6338959  
; GENERAL INFORMATION:  
; APPLICANT: HATADA, Yuji  
; APPLICANT: IGARASHI, Kazuaki  
; APPLICANT: OZAKI, Katsuya  
; APPLICANT: ARA, Katsutoshi  
; APPLICANT: KAWAI, Shuji  
; APPLICANT: ITO, Susumu  
; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PULLULANASE AND  
; TITLE OF INVENTION: ALKALINE ALPHA-AMYLASE ACTIVITIES  
; FILE REFERENCE: 2173-105P  
; CURRENT APPLICATION NUMBER: US/09/514.302  
; CURRENT FILING DATE: 2000-02-28  
; EARLIER APPLICATION NUMBER: 08/952,084  
; EARLIER FILING DATE: 1997-11-10  
; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 833  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-514-302-3  
Query Match 2.8%; Score 102.5; DB 4; Length 833;  
Best Local Similarity 19.0%; Pred. No. 1.6;  
Matches 108; Conservative 71; Mismatches 199; Indels 191; Gaps 25;  
Qy 231 VYAGADNRPA-----EYSKDNKPKYKPVYFAAVSMOGYKAD-DYAMTIGFPGSTDRYLTSMG 285  
Db 238 VTAGVKNLPIITAIDFEGNRHEGSATLEVOARTITGEKADFQDQSVVYFMLTDREF----- 353  
Qy 286 VEDRIENENNR-----IEVRGIKQ-----GIWKEAMSADQATRI 320  
Db 354 --DGSSNNNDPHGIGYDTSGTGVGGDFKGTORLDYDELGINTIWISPV-----VDNI 407  
Qy 321 KYASKYAQSAN-----YMKNSIG-----MNRGL-ARLDVI--- 349  
Db 408 KFDVHRHSEGPDPYAYHYGNADNFGELNPHFGSMADPHEMIDAHERGIIKIMVDVVLNH 467  
Qy 350 ---GRKRA-----EERA-FADWIRKNGSAVYGDVLSLEKAYKEGAKANREMT 394  
Db 468 TCYGLKPGDSSSVANFPPTDEDRARFDGMLRDGGSGEVRGE-LAGLPDFLTENPDVREQVV 526  
Qy 395 Y-----LSETLFGT-----EVVRFQAQANALATNPDAHAGILKSLDDKYYK 435  
Db 527 QWQTDWIEKSRTAGNTIDYFRVDTVKHVEDTWMFAFNKALTAMPEHKLIGEAMGANVN 586  
Qy 436 DYLPSLDKRVLPAMDIVRRRIPADKLPDIFKNVIDKFKGDTKYADFVDFKSVVPYSD 495  
Db 587 DDLGVLSNGMDSLDF-----DFKNYARDFANGQLDAYQQ 622  
Qy 496 KFHA-----MLKSMDEKFAKAEIKDPDAVELSKSVIAARAIOADAMANAYA 542  
Db 623 KLEARNKLNNTATLGOFLGSHDEDFEVEVEGD---LGKYVQAASLQLTA----- 670  
Qy 543 IEKRRKRFAGLREMPGRALPSDANFTMRMSYGSIKGYEPDQGAWYNYHTTGKGVLEKQ 602  
Db 671 --KQPVIIYGEELGPKG---ND-----YPYTNRQN-MPMD 702  
Qy 603 DPKSDEFVAQENILDLFRTKNYGRYAENGQHLIAFLSNNDITGNGSGSPVFDKNGRLIGL 662  
Db 703 DVDGNEILHEYQKLLAFRNDNPNTFAKGRKVA---GSDSEGYLFLFSRTYGENSVVGL 759  
Qy 663 AFDGNWEAMSGDIEFEPDLQRTISVDIRY 691  
Db 760 ----NTEAAAKDVTILNFGSSEAVVTD-RY 783

RESULT 14  
US-09-134-001C-5438  
; Sequence 5438, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134.001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5438  
; LENGTH: 856  
; TYPE: PRT  
; ORGANISM: staphylococcus epidermidis  
US-09-134-001C-5438

[illegible]

```

RESULT 15
US-09-134-001C-3032
; Sequence 3032, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3032
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3032

```

Query Match

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2003, 15:31:51 ; Search time 52.8284 Seconds  
(without alignments)  
1795.900 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MQMKLSILGALLLGASG.....LFMIDKWCQPRLIQELKLI 712

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3719	100.0	712	23	AAO15205
2	274	7.4	52	23	AAO15206
3	146.5	3.9	1368	23	ABP27518
4	133	3.6	883	22	AAU37789
5	130	3.5	883	20	AAU37789
6	130	3.5	883	21	AAU90514
7	130	3.5	883	23	AAU97883
8	127.5	3.4	1194	18	AAW21725
9	127	3.4	883	22	AAU38091
10	126.5	3.4	1370	23	ABP27517
					Porphyromonas ging
					Porphyromonas ging
					Streptococcus poly
					Streptococcus pneu
					S. pneumoniae vals
					Streptococcus pneu
					Streptococcus pneu
					Modified streptoki
					Streptococcus pneu
					Streptococcus poly

11	123.5	3.3	713	20	AAU43380	M. catarrhalis str
12	122.5	3.3	26	23	AAO15221	Porphyromonas ging
13	122.5	3.3	800	18	AAW21723	Plasminogen-bindin
14	122.5	3.3	813	18	AAW21728	Wild type plasmino
15	122.5	3.3	1181	18	AAW21727	Streptokinase/malt
16	122.5	3.3	1194	18	AAW21724	Modified streptoki
17	122.5	3.3	1194	18	AAW21726	Streptokinase/malt
18	118	3.2	882	23	ABP25774	Streptococcus poly
19	117.5	3.2	951	20	AAU34536	Porphyromonas ging
20	117.5	3.2	953	20	AAU34403	Porphyromonas ging
21	115.5	3.1	2274	23	ABP30377	Streptococcus poly
22	115.5	3.1	2278	23	ABP28340	Streptococcus poly
23	114.5	3.1	657	23	AAO17805	H influenzae BVH-N
24	114	3.1	711	17	AAU88649	Neisseria meningit
25	113	3.0	708	21	AAU43567	Human cancer assoc
26	113	3.0	883	23	ABP25773	Streptococcus poly
27	113	3.0	2042	19	AAU56319	Haemophilus paraga
28	112.5	3.0	907	22	AAU78650	XRN-100. Unidenti
29	112	3.0	724	22	ABG15384	Novel human diagno
30	112	3.0	724	22	ABG17531	Novel human diagno
31	110.5	3.0	1027	14	AAU42203	Protein L. Peptoc
32	110.5	3.0	1027	14	AAU43699	Protein L. Peptoc
33	110	3.0	477	23	ABU93927	Herbicidally activ
34	110	3.0	705	17	AAU88645	Neisseria meningit
35	109.5	2.9	845	21	AAU90942	Cenarchaeum symbio
36	109	2.9	416	21	AAU34433	Gene 46 human secr
37	109	2.9	416	21	AAU34433	Human secreted pro
38	109	2.9	721	23	ABU47955	Listeria monocytlog
39	109	2.9	1577	17	AAU91047	Alpha-D-glucosyltr
40	109	2.9	1640	23	ABU54727	Lactococcus lactis
41	109	2.9	2234	21	AAU81502	Streptococcus pneu
42	108.5	2.9	950	21	AAU42742	Human ORF2506
43	108.5	2.9	950	22	AAU93087	Human protein sequ
44	108	2.9	2027	22	ABG07898	Novel human diagno
45	108	2.9	4536	20	AAU96826	Amino acid sequenc

ALIGNMENTS

RESULT 1  
AAO15205  
ID AAO15205 standard; Protein; 712 AA.  
XX  
AC AAO15205;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).  
XX  
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
KW periodontitis.  
XX  
OS Porphyromonas gingivalis.  
XX  
PN WO200238742-A2.  
XX  
PD 16-MAY-2002.  
XX  
PF 08-NOV-2001; 2001WO-US46782.  
XX  
PR 08-NOV-2000; 2000US-246827P.  
XX  
PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
PI Travis J, Potempa JS, Banbula A, Bugno M;  
XX  
DR WPI; 2002-490075/52.  
XX  
DR N-PSDB; AAL43635.  
XX  
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
the dipeptidylpeptidase for protecting an animal from periodontal



KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus pyogenes.

PN WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tetelin H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN68149.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX Claim 1; Page 3578; 4525pp; English.

PS The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (I) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

XX Sequence 1368 AA;

Query Match 3.9%; Score 146.5; DB 23; Length 1368;  
Best Local Similarity 19.5%; Pred. No. 0.004;  
Matches 167; Conservative 105; Mismatches 228; Indels 357; Gaps 43;

QY 35 NQENLDRMRELGFILPLDLSYFDPKPSANAVIPGGCTGTVSDQGL--IFTNHHGCGY 92  
DB 609 NEENEDILEDIVLTLTLE---DREMEERLKYAH-----LFDDKVMKQLKRRYRTGW 659  
QY 93 GAI-----QSQSTVDHDLR--DGFVSR-----TMGEELPIPGLS---- 125  
DB 660 GLSKRLKINGIRDKSGGTI--LDFLKSDFANRPMQLIHDSLTFFKEDIQAQVSGGCD 718  
QY 126 -----VKYLRIVKVD---KVEGO-----LKGITDEM 150  
DB 719 SLHEHIANLAGSPAIIKGILOTQVKVVDLVKVMGRHKPENIVEMARENQTTQKQKNSR 778  
QY 151 ERLKRAQVQELAK---KENADEN-QLCIVEPFSNNEYELIVY-----DVFKDVRWVF 201  
DB 779 ERMKRIEIGIKELSGQILKEHPVENTQL-----QNEKLYLYLQNGROMYVDQELDI 830  
QY 202 APPSSVGFGGDTDNWMPRHTGDSFVRVAGADNRPAEYSKDNKPKYKPVYFAVSMQG 261  
DB 831 -----NRLSDYDVH-----IVPQS 845

QY 262 YKADYAMTIGPGSTDRYLTSWGVEDRIENENNRIEVRGIKQIGWKEAMSADQATRIK 321  
DB 846 FLKDD-----SIDNKVLTRSDKNGKSDNVPSEEVKKMKNYWQLLNAKLITQRK 896  
QY 322 YASKYAQSANYWKNSIGNRGLARLDVIG---RKRAEERAFADWI-----RKNGK---- 368  
DB 897 F-----DNLTKAERG---GLSELDKAGFTKRQLVETROITKHVAQILDSRMNTKYDEN 946  
QY 369 -----SAVYGDVL-----SSLEK 381  
DB 947 DKLIREVKVITLKSCLVSDFRKDFOFYKVRNINNYHHAHAYLNAVVGITALIKKPKLES 1006  
QY 382 AYKEG-----AKANREMTYLSLTFGGTEVVRFAQFANALATNPDAHAGILKS- 429  
DB 1007 EFVYGDYKVDVVRKMIKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN 1066  
QY 430 -----LDDKYDYLPISLDRKVLPA-AMLDIVRRR-----IP---ADKLDPDIFK 467  
DB 1067 GETGEIVWDKGRDF--ATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKL----- 1119  
QY 468 NVIDKKFKGDTKKYADFVFDKSVVPYSDKFHA-----MLKS-----MDKEKFA 510  
DB 1120 --IARKKWDPKKYCG--FDSPTVAYSVLVYAKVEKSKKLKSVKELLGITIMERSSE 1175  
QY 511 -----KATEKPAVELSKSVIAAARAIQADAMANAAYAEKGRKFLFAGLREMY 558  
DB 1176 KNPIDFLEAKGYKEVKDLIIKLPR-----YSLFELENGRKRMLASAGELQ 1221  
QY 559 PGR--ALPSPD--ANFTMRMS--YCSIKGYEPDQG---ANYNYHTTGKGVLEKQDPKSDFA 610  
DB 1222 KGNELALPSKYVNFYLYLASHVEKLGK--SPEDNEQKQLFVEQH---KHVLEDEIEQISEFS 1277  
QY 611 VQ-----ENTLDLFRTKNYG-----RYA 628  
DB 1278 KRVILADANLKVLSAYNKHDKPIREQAENIIHLTLTLNLCAPAPAFKYFDTIDRKRYT 1337  
QY 629 ENGQHLIAFLSNNDITG 645  
DB 1338 STKEVLDATLIHQISITG 1354  
RESULT 4  
AAU37789  
ID AU37789 standard; Protein; 883 AA.  
XX AC AU37789;  
XX DT 14-FEB-2002 (first entry)  
XX DE Streptococcus pneumoniae cellular proliferation protein #218.  
XX KW Antisense; prokaryotic cellular proliferation protein;  
XX KW Antibiotic; antibacterial; drug design.  
XX OS Streptococcus pneumoniae.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
XX PR 23-MAY-2000; 2000US-206848P.  
XX PR 26-MAY-2000; 2000US-207272P.  
XX PR 23-OCT-2000; 2000US-242578P.  
XX PR 27-NOV-2000; 2000US-253625P.  
XX PR 22-DEC-2000; 2000US-257931P.  
XX PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;



QY 536 AMANAYAIKGRKRLFFAGLREMYPGRALPS-----DANFTMRMSYSGIKGYEPQDG 586  
Db 372 AIANQDTEDK-----VEEYPPRFNDTFLQWENHVDWVISRLQWGH-----QIP 416  
QY 587 AWNYHTTGKGVLEKODPKSDEFVAQENILLD 617  
Db 417 AWYN--ADGEMTVGEEAPEGQDQEDVLD 445  
RESULT 6  
AAU90514  
ID AAU90514 standard; Protein; 883 AA.  
XX  
AC AAU90514;  
XX  
DT 15-AUG-2000 (first entry)  
XX  
DE Streptococcus pneumoniae valyl tRNA synthetase (vals).  
XX  
KW Valyl tRNA synthetase; vals; inhibitor; stringent response;  
KW drug screening; antibacterial; antibiotic; genetic immunisation;  
KW antibody; bacterial infection; meningitis.  
XX  
OS Streptococcus pneumoniae strain 0100993.  
XX  
PN US6051413-A.  
XX  
PD 18-APR-2000.  
XX  
PF 24-SEP-1998; 98US-0159539.  
XX  
PR 18-APR-1996; 96GB-0007791.  
PR 17-OCT-1997; 97US-0953492.  
PR 18-APR-1997; 97US-0844064.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
PI Brown JR, Lawlor EJ, Wang M, Jaworski DD;  
XX  
DR WPI; 2000-338311/29.  
DR N-PSDB; AAA14369.  
XX  
PT Novel vals polypeptides of valyl tRNA synthetase family useful for  
PT treating otitis media, conjunctivitis, pneumonia and bacteremia  
PT comprises a specified amino acid sequence -  
XX  
PS Claim 5; Columns 9-10; 21pp; English.  
XX  
CC This sequence represents Streptococcus pneumoniae valyl tRNA synthetase  
CC (vals, NCIMB Deposit No. 40794). Streptococci are the cause of several  
CC types of human diseases, including otitis media, conjunctivitis,  
CC pneumonia, bacteraemia, sinusitis, pleural empyema, endocarditis and  
CC especially meningitis. The frequency of S. pneumoniae infections has  
CC risen dramatically over the past 20 years, probably due to the emergence  
CC of multiply antibiotic resistant strains and an increasing population of  
CC immunocompromised people. Vals represents a target for new antibacterial  
CC agents. Inhibition of tRNA synthetases such as vals leads to a reduction  
CC in the levels of charged tRNA, which triggers a cascade of responses  
CC (known as the stringent response) resulting in a state of dormancy in the  
CC bacterium. Vals, its variants and fragments, anti-vals antibodies, vals  
CC inhibitors and nucleotides encoding vals may be used in the diagnosis,  
CC prevention and treatment of bacterial infections such as meningitis.  
CC Vals can be used to screen compounds for inhibitory activity. Vals  
CC may also be useful as an antigen for vaccination of a host to produce  
CC specific antibodies which protect against bacterial invasion into  
CC damaged tissues. Such antibodies could, for example, prevent the  
CC adherence of bacteria to wounds. Nucleotides encoding vals may be used as  
CC diagnostic reagents and therapeutic or prophylactic agents, particularly  
CC for genetic immunisation.  
XX  
SQ Sequence 883 AA;

Query Match 3.5%; Score 130; DB 21; Length 883;

Best Local Similarity 20.4%; Pred. No. 0.055;  
Matches 104; Conservative 63; Mismatches 168; Indels 176; Gaps 24;  
QY 192 DVF-----KDVRFVAPPSSVGKFGGDTDNWMPRHRTGDFSVFRVYAGADNRPAYS 243  
Db 26 DVFKSGQKAKPYSIVIPPNVTGK-----HLG-----HA 57  
QY 244 KDNKPKPVYFAAVSMOGYKADDYAMTIFGPGSTDRYL-TSMGVEDRIENENNPRIYRG 302  
Db 58 WDT-TLQDIIIRQKRMQGF-----TLWLPQMDHAGIATQAKVEERLGRGEGISRYDLGR 110  
QY 303 -----IKQGIWKEAMSADQATRIKYA-SKYAQSANTYKWSIGNRGLARLDVIGRKRAER 357  
Db 111 ESFLTKVWEKD-----EYATTIKEQMGKGLSDVYSRERFTLDEGLS-----KAVRK 158  
QY 358 AFADWIRK-----NGKSAVYGD-----VLSLEKAYKEGAKANREMTYL----- 396  
Db 159 VFVDLYKKGWYRGFEIINWDPAAFTALSDIEVIHKDVEGAFYHMYMLEDSRVLEVAT 218  
QY 397 --SETLFGGTVEVRFAPANALATNPDAHAGILKSLDDKYDYLPs-----LDRKVLPA 449  
Db 219 TRPETMFGDV-----AVAVNPE-----DPRYKDLICKNVILPIANKLPIV 259  
QY 450 LD-----IVRRRTIPADKLPIDFIKNVIDKKFKGDTKKYADVDFKSVVVPYSDK 496  
Db 260 GDEHADPEFGTGVWKITPAHDPNDFLVGQRHNLQVNVNMDGTMNDLAFEFSGM---DR 316  
QY 497 FHMLKSMDEKFAKAEKOPAVELSKSVIAARA-----IQAD 535  
Db 317 FEAR-----KAVVAKLEEIGALVKIEKRVHSGHSGERTGVVVEPRLSQWFMVKMDOLAKN 371  
QY 536 AMANAYAIKGRKRLFFAGLREMYPGRALPS-----DANFTMRMSYSGIKGYEPQDG 586  
Db 372 AIANQDTEDK-----VEEYPPRFNDTFLQWENHVDWVISRLQWGH-----QIP 416  
QY 587 AWNYHTTGKGVLEKODPKSDEFVAQENILLD 617  
Db 417 AWYN--ADGEMTVGEEAPEGQDQEDVLD 445  
RESULT 7  
AAU97883  
ID AAU97883 standard; Protein; 883 AA.  
XX  
AC AAU97883;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Streptococcus pneumoniae Vals.  
XX  
KW Vals; microbial infection; antibacterial.  
XX  
OS Streptococcus pneumoniae.  
XX  
PD JP2002119292-A.  
XX  
PF 17-OCT-1997; 2001JP-0216648.  
XX  
PR 17-OCT-1997; 97JP-0321886.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX  
DR WPI; 2002-475332/51.  
DR N-PSDB; ABK52414.  
XX  
PT Novel vals polynucleotide useful for producing a vals polypeptide, used  
PT in the treatment of vals-related disorders -  
XX  
PS Claim 1; Fig 4; 24pp; Japanese.  
XX  
CC The invention describes an isolated vals polynucleotide encoding a mature





Db 895 GKVPADKGS-VTLTPQVQEFLLSGHVRVRYKE-----KPIQNOA--KSDVVEYTV 944  
 Qy 435 -----KDYLPSSL-DRKVLPAMLDIIVRRIPADKLPDIFKNVIDKKFKGDTKKYADFV 485  
 Db 945 QFTPLPDDFRFGKDKYLLKTL--AIGDITISQELLAQAQSILNKHNP-----YTIYE 998  
 Qy 486 FDKSVYPYSDKFAHMLKSMKDEKFAKAEKDPFAVELSKSVIAAARAIAQADAMANA-YAIE 544  
 Db 999 RDSSIVTHNDIFRILPMDQEFYVRVNRQAYRINKKSGLNBEINNTDLISEKYYVLK 1058  
 Qy 545 KKK-----RLFAGLREMYPGRALPSDANFT---NRMSYGSIKGYEPQDGA--W 588  
 Db 1059 GKEKPYDPFDRSHLKLFTIKYVVDVNTNELLKSEQLLTASERNLDFROL--YDPRDKAKLL 1116  
 Qy 589 YN-----YHTFGKGVLEKQDKPSDEFAVOENILDIFR--TKNYGR--YAENGQLHI 635  
 Db 1117 YNLDFAFGIMDYTLTK-----VEDNHDDTNRITVYMGKRPGEGENASTYHL 1162  
 Qy 636 AF 637  
 Db 1163 AY 1164  
 RESULT 9  
 AAU38091  
 ID AAU38091 standard; Protein; 883 AA.  
 AC AAU38091;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 STreptococcus pneumoniae cellular proliferation protein #520.  
 DE  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI: 2001-611495/70.  
 DR N-PSDB: AAS55950.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 13684; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 883 AA:  
 Query Match 3.4%; Score 127; DB 22; Length 883;  
 Best Local Similarity 20.2%; Pred. No. 0.1; Mismatches 167; Gaps 24;  
 Matches 103; Conservative 65;  
 Qy 192 DVF-----KDRMVFAPPSSVGKFGGDTDNMMWPRHTGDSVFRVYAGADNRPAEYS 243  
 Db 26 DVKPSGQKAKPYSIVIPPNVTCKL-----HLG-----HA 57  
 Qy 244 KDNPKYKPVYFAVSMQGYKADYAMTIGFPGSTDRYL-TSWGVEDRIENENNPRIEVRG 302  
 Db 58 WDT-TLQDIIIRQKRMOGFD-----TLWLPGMDHAGIATQAKVEERLGRGEGITRYDLGR 110  
 Qy 303 ---IKQIWKEAMSADQATRIKYA-SKYAOSANYWKNISGNRGLARLDVIGRKRAEER 357  
 Db 111 ESFLTKVWENK-----EYATTIKEQMGKMGSLVDYSRERFTLDEGLS-----KAVRK 158  
 Qy 358 AFADWIRK-----NGKSAVYGD-----VLSSLEKAYK-----EGAKANREMT 394  
 Db 159 VFDVLYKKGWYRGFEIINWDPAAARTALSDIEVIHKVEGAFYHMYMLEDSRALEVAT 218  
 Qy 395 YLSETLFGGTEVYRFAQFANALATNPDAHAGILKSLDDKYKDYLP-----LDRKVLPA 449  
 Db 219 TRPETMFGDV-----AVAVNPE-----DPYKDLIGKNVILPIANKLPIV 259  
 Qy 450 LD-----IVRRRIPADKLPDIFKNVIDKFKGDTKKYADFVFKSVVVPYSDK 496  
 Db 260 GDEHADPEFGGVVKITPAHDNDPDLVGRHNLPOVNVNDDGTWNLAFESGM---DR 316  
 Qy 497 FHMLKSMDEKFAKAEKDPFAVELSKSVIAAARA-----IQAD 535  
 Db 317 FEAR-----KAVVAKLEEIGALVKIEKRVHSGHSGERTGVVVEPRLSQWFVKMQDLAKN 371  
 Qy 536 AMANAYAIKKGRLFFAGLREMYPGRALPS-----DANFTMRMSYGSIKGYEPQDG 586  
 Db 372 AIANQDTEDK-----VEFYPPRFNDTFLQWMENVDWISRLQMWGH-----QIP 416  
 Qy 587 AWYNYHTTCKGVLEKQDKPSDEFAVOENILD 617  
 Db 417 AWYN--ADGEMVYGEAPEGGGTQDEDDVLD 445  
 RESULT 10  
 ABP27517  
 ID ABP27517 standard; Protein; 1370 AA.  
 XX  
 AC ABP27517;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 4210.  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.

XX PD 02-MAY-2002.  
 XX PF  
 XX PR 29-OCT-2001; 2001WO-GB04789.  
 XX PR 27-OCT-2000; 2000GB-0026333.  
 XX PR 24-NOV-2000; 2000GB-0028727.  
 XX PR 07-MAR-2001; 2001GB-0005640.  
 XX PR (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 XX Tettelin H;  
 XX WPI: 2002-352536/38.  
 XX N-PSDB; ABN68148.  
 XX  
 XX New Streptococcus protein for the treatment or prevention of infection  
 XX or disease caused by Streptococcus bacteria, such as meningitis, and  
 XX for detecting a compound that binds to the protein -  
 XX  
 XX Claim 1; Page 3577-3578; 4525pp; English.  
 XX  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 XX the specification. The proteins have antibacterial and antiinflammatory  
 XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 XX antibodies that bind (I) are used in the manufacture of medicaments for  
 XX the treatment or prevention of infection-or disease caused by  
 XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 XX Nucleic acids encoding (I) are used to detect Streptococcus in a  
 XX biological sample. (I) is used to determine whether a compound binds to  
 XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 XX used as a vaccine or diagnostic composition. The disease caused by  
 XX Streptococcus that is prevented or treated may be meningitis. Nucleic  
 XX acid encoding (I) may be used to recombinantly produce (I) and may be  
 XX used in gene therapy. Antibodies to (I) are used for affinity  
 XX chromatography, immunoassays, and distinguishing/identifying  
 XX Streptococcus proteins.  
 XX  
 XX Sequence 1370 AA;  
 XX  
 XX Query Match 3.4%; Score 126.5; DB 23; Length 1370;  
 XX Best Local Similarity 19.8%; Pred. No. 0.22;  
 XX Matches 163; Conservative 111; Mismatches 260; Indels 289; Gaps 43;  
 XX  
 XX QY 35 NQENLDRMRELGLTLPDLSYFDPKPSIANAVVIFGGCTGTVSDQGLIFTNHHCYGA 94  
 XX DB 611 NPDNIEILEDIVQTLTFEDREMIRKRLNRYKDLF-----TESQLKLYRRHYTGWR 663  
 XX QY 95 I-----QSQSTVDHDLRDCGFVSRWGEELPIPLSVKYLKIVKV-----TD 137  
 XX DB 664 LSAKLINGIRDESQKTLIDLDGSRNRNFMQLINDGLGFSKTSISKAQAGSHSDNLK 723  
 XX QY 138 KVEGOLKG-----ITDNERLR--KAQVCOELAKKENADENQ----- 173  
 XX DB 724 EVVAGELAGSPAIGKGLQSLKIVDELVKVNGVEPEQIVVEMA-RENQTTNQRGRNRQRY 782  
 XX QY 174 -----LC-----IVEPPYSNNE-----YF-----LIVYD 192  
 XX DB 783 KLDDGVKNLADNLNGILNKEYPTDNLQALQNERLFYLYLQNGRDMYTGALDIDNLSQVD 842  
 XX QY 193 VFQDVRMVFAPSSVKGFGDDTNMWRPRHTGDFSVFRVYAGADNR-----PAEYSKD 245  
 XX DB 843 IDHIIQPAFIKDSI-----DNRV-----LVSSAKNRKSDDDVPSLDIVKD 883  
 XX QY 246 NKPY-KPVYFAAVSMQ-----GYKADDVAMTIGFPGSTDRYLTSMG---VEDR 289  
 XX DB 884 CKVFWKKLLDAKLMRSQRKYDNLTKAERGGLTSDDKAREIQROLVETROITKHHVRLDER 943  
 XX QY 290 IENENNP-----RIVRGIKOGINKEAMSADQATRIKYASKYASANYKWKNSIGMNRGL 343

Db 944 FNNELDSKGRRIKRVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVAKAIL 1003  
 QY 344 ARLDVIGRKRAREERAFADWIRKNGKSAYVGDVLSLEKAYKEGAKANREMTYLSE----- 398  
 Db 1004 TRYQQL-----EPEF-----VYGDY--PKYNSYKTRKSATERKLFYFSNIMNPF 1044  
 QY 399 ----TLFGGTEVVRFAQAFANALATNPDAHAGILKSLDDKYKDYLPSLDRKVL--PAMLDI 452  
 Db 1045 KTKVTLADGTVVVK-----DDIEVNNDTG-----EIVWDKKHF--ATVRKVLSPONNIV 1093  
 QY 453 VRRRIP-----ADKLDPIDFNVIDKKFKG---DTKKYADVFDFKSVVPYSD 495  
 Db 1094 KKTEIQGTGFSKESILAHGNSDKL-----IPRKTDIYLDPKKYGG--FDSPIVAYSV 1144  
 QY 496 KFHMLKS-----MDKEKFAKATEKDPAVEL--SKSVTAARAQAADAM 537  
 Db 1145 LVVADIKKGAOKAKLKTIVTELLGITIMERSRF-----EKNPSAFLESKYL-----NIRADKL 1196  
 QY 538 A-----NAVAIKKGKRLFFAGLREMYVGRALPDSANFT----MRKSYGSIKGYEPQD--- 585  
 Db 1197 IILPKYSLELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKG-KPEEIEK 1255  
 QY 586 -GAWNYHTT-----GKGV-----LEK-----QDPKS-----DEFAVOENILD 617  
 Db 1256 KOEFVNOHVSYFDDILQLINDESKRVILADANLEKINKLYQDNKENISVDELA--NNIIN 1313  
 QY 618 LFRTRNYG-----RVAENCOLHIAFLSNNDING 645  
 Db 1314 LFTFTSLGAPAAKFFDKIVDRKRYTSTKVELNSTLIHQISITG 1356  
 XX  
 XX RESULT 11  
 XX AAY43380  
 XX ID AAY43380 standard; Protein: 713 AA.  
 XX AC AAY43380;  
 XX DT 26-JAN-2000 (first entry)  
 XX DE M. catarrhalis strain LES1 tbp2 protein.  
 XX  
 XX TbpB gene; Tbp2; transferrin binding protein; diagnosis; otitis media;  
 XX genetic immunisation; Moraxella infection; antigen; vaccine; detection;  
 XX antitumour antibody production; therapy.  
 XX  
 XX OS Moraxella catarrhalis.  
 XX  
 XX XN WO9952947-A2.  
 XX PD 21-OCT-1999.  
 XX PF 12-APR-1999; 99WO-CA00307.  
 XX PR 14-APR-1998; 98US-0059584.  
 XX PA (CONN-) CONNAUGHT LAB LTD.  
 XX PI Myers LE, Schryvers AB, Harkness RE, Loosmore SM, Du R, Yang Y;  
 XX PI Klein MH;  
 XX DR WPI: 1999-620376/53.  
 XX DR N-PSDB; AAZ31948.  
 XX  
 XX Nucleic acid encoding transferrin binding protein 2 of Moraxella  
 XX catarrhalis, useful for diagnostics, immunization and recombinant  
 XX protein production -  
 XX  
 XX Claim 7; Fig 6; 114pp; English.  
 XX  
 XX This sequence is the Moraxella catarrhalis strain LES1 transferrin  
 XX binding protein (Tbp2) of the invention. The DNA sequence is also  
 XX referred to as the TbpB gene. The TbpB gene is used to produce





Db 445 RPAHGGKTEQGLSPKSPFAFATDSGAMSHKLEKADLL--KAIQEQLIAN-VHSNDDYFEVI 501  
Qy 367 --GRSAVYGDVLSLEKAYKEGAKANREMTYLSLTFGGTEVVRFAOFANALATNPDAHA 424  
Db 502 DFASDATITDRNGKVFADKDG--VTLPQVQFELLSGHVRYKE-----KPTQNG 553  
Qy 425 GILKSLDDKY-----KDYLP--DRKVLPAMLDIVRRRIPADKLPDIFKNVIDKK 473  
Db 554 A--KSDVVEYTVQFTPLNPDODDFRGLKDTLLKTL--AIGDTITTSQELLAQAOSILNKN 609  
Qy 474 FKGDTKKYADVFDKSVVPYSDKPHAMLSMDKEKFAKAEKDPAVELSKSVIAARAIO 533  
Db 610 HPG-----YTIYERDSSIVTHNDIFRTILPMDQEFYRVKNREQAYRINKSKSGLNEEINN 665  
Qy 534 ADAMANA-YAIEKKG-----RLFFAGLREMYPGRALPSDANFT---MRMSYCSI 578  
Db 666 TDLISEKYVVLKKGKPYDPFDRSHLKLFTIKYVDVDTNELLKSEQLLTASERNLDFRDL 725  
Qy 579 KGYEPQDGA--WYN-----YHTTGKGVLEKODPKSDEFVAOENILDLFR--TKNYG 625  
Db 726 --YDPRKAKLLYNLDAFGIMDYTLTK-----VEDNHDNTNRIITVYMG 769  
Qy 626 R--YAENGOLHIAF 637  
Db 770 KRPEGENASYHLAY 783  
RESULT 15  
AAW21727  
ID AAW21727 standard; protein; 1181 AA.  
XX AC AAW21727;  
XX DT 01-OCT-1997 (first entry)  
XX DE Streptokinase/maltose binding protein fusion protein, rSKdelta14.  
XX KW Plasminogen-binding fragment; streptokinase; degradation; MBP;  
XX KW thrombolytic agent; blood clot; bolus; maltose-binding protein.  
XX OS Streptococcus equisimilis.  
XX FH Key Location/Qualifiers  
FT Protein 1..381  
FT /label= Maltose binding protein  
FT /note= "acts as blocking group"  
FT Protein 382..1181  
FT /label= Modified Streptokinase  
FT /note= "Has N-terminal 14 amino acids deleted"  
XX WO96411883-A1.  
XX PN 27-DEC-1996.  
XX PD 07-JUN-1996; 96WO-US09640.  
XX PF 09-JUN-1995; 95US-0488940.  
XX PR (HARD ) HARVARD COLLEGE.  
XX PA Reed GL;  
XX PI WPI; 1997-065469/06.  
XX DR Modified forms of streptokinase resistant to enzymatic cleavage -  
XX PT useful as thrombolytic agents in treating thrombosis and in medical  
XX PT equipment  
XX PS Example 1; Page 12; 65pp; English.  
XX CC This sequence represents a fusion protein between maltose-binding  
CC protein and a mutant form of the plasminogen-binding fragment

CC of streptokinase which has the N-terminal 14 amino acids deleted.  
CC This fusion protein was used in the design of a modified streptokinase  
CC has an in vitro degradation rate at least 2 times slower than that of  
CC native streptokinase. Compounds containing modified streptokinases  
CC are specifically used as thrombolytic agents for dissolving blood  
CC clots in vivo in a mammal, preferably at a dose of 20000 U/kg,  
CC optionally as a bolus rather than by continuous infusion.  
XX SQ Sequence 1181 AA;  
Query Match 3.3%; Score 122.5; DB 18; Length 1181;  
Best Local Similarity 19.3%; Pred. No. 0.39;  
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;  
Qy 75 GITVSDQGL-----IFTNHHCGYGAIQSQSVTDHLYLRDGFVSRVTGMEELPIP---GLS 125  
Db 558 GDTTTSQELLAQAOSILNKNHPGVIYERDSSI---VTHNDNIDFTI---LPMQOEFTYR 611  
Qy 126 VKYLRKIVKVTDKVEGQLGITDEMERLKAQEVCOELAKKE----- 167  
Db 612 VKNREQAYRINKK-----SGLNEEINNTDLISEKYVVLKKGKPYDPFDRSHLKLFTIKY 666  
Qy 168 -NADENQLCIVEPYSNNEY---FLIVYDVFKDVMYFAPSPSSVKKFC-----GUTDN 216  
Db 667 VVDVTNELLKSEQLLTASERNLDFRDLYDPRDKAKLLY---NNLDAFGIMDYTLTKVED 723  
Qy 217 WMPRHTGDFSFRVYAGADNRPAEYSKDNKPKYKPVYFAAVSMQGYKADYAMTIGPGS 276  
Db 724 ---NHDDTNRIITVYMG--KRP---EGENASYHLAY-----D 752  
Qy 277 TDY-----LTSWGVEDRIENENPNRI-----EVGKIGKIGWKEAMSAOAT 318  
Db 753 KDRVTEEREVYSYLYRTGTPIPNPNKNSQLVSVAGTVEGTNODISLKFFEIDLTS 812  
Qy 319 RIKYASKYAO-----SANYWKNISGMNRLARLDVIGRKRAEERAFADWIRN----- 366  
Db 813 RPAHGGKTEQGLSPKSPFAFATDSGAMSHKLEKADLL--KAIQEQLIAN-VHSNDDYFEVI 869  
Qy 367 --GRSAVYGDVLSLEKAYKEGAKANREMTYLSLTFGGTEVVRFAOFANALATNPDAHA 424  
Db 870 DFASDATITDRNGKVFADKDG--VTLPQVQFELLSGHVRYKE-----KPTQNG 921  
Qy 425 GILKSLDDKY-----KDYLP--DRKVLPAMLDIVRRRIPADKLPDIFKNVIDKK 473  
Db 922 A--KSDVVEYTVQFTPLNPDODDFRGLKDTLLKTL--AIGDTITTSQELLAQAOSILNKN 977  
Qy 474 FKGDTKKYADVFDKSVVPYSDKPHAMLSMDKEKFAKAEKDPAVELSKSVIAARAIO 533  
Db 978 HPG-----YTIYERDSSIVTHNDIFRTILPMDQEFYRVKNREQAYRINKSKSGLNEEINN 1033  
Qy 534 ADAMANA-YAIEKKG-----RLFFAGLREMYPGRALPSDANFT---MRMSYCSI 578  
Db 1034 TDLISEKYVVLKKGKPYDPFDRSHLKLFTIKYVDVDTNELLKSEQLLTASERNLDFRDL 1093  
Qy 579 KGYEPQDGA--WYN-----YHTTGKGVLEKODPKSDEFVAOENILDLFR--TKNYG 625  
Db 1094 --YDPRKAKLLYNLDAFGIMDYTLTK-----VEDNHDNTNRIITVYMG 1137  
Qy 626 R--YAENGOLHIAF 637  
Db 1138 KRPEGENASYHLAY 1151

Search completed: May 25, 2003, 15:33:52  
Job time : 72.8284 secs



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:47:08 ; Search time 64.7273 seconds  
(without alignments)  
1060.959 Million cell updates/sec

Title: US-10-008-355-2  
Perfect score: 3719  
Sequence: 1 MQMKLSILLGALLLGASG.....LFMDKWGCPRLIQELKLI 712

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3719	100.0	712	9	US-10-008-355-2
2	3652	98.2	699	9	US-10-008-355-8
3	1276	34.3	720	9	US-10-008-355-9
4	907.5	24.4	716	9	US-10-008-355-7
5	883.5	23.8	734	9	US-10-008-355-5
6	807.5	21.7	732	9	US-10-008-355-6
7	274	7.4	52	9	US-10-008-355-3
8	133	3.6	883	10	US-09-815-242-13382
9	127	3.4	883	10	US-09-815-242-13684
10	122.5	3.3	26	9	US-10-008-355-18
11	113	3.0	708	10	US-09-925-301-1012
12	110	3.0	2042	9	US-10-192-584-6
13	109.5	2.9	845	9	US-10-027-806-62
14	109.5	2.9	845	9	US-10-034-623-62
15	109.5	2.9	845	9	US-10-027-801-62
16	109	2.9	687	9	US-09-969-384-27
17	106.5	2.9	523	10	US-09-815-242-11918
18	106	2.9	972	9	US-10-100-049-23
19	105	2.8	1023	9	US-09-884-696-5

20	104	2.8	1879	9	US-09-971-536-70	Sequence 70, Appl
21	104	2.8	4563	9	US-09-802-640-32	Sequence 32, Appl
22	103.5	2.8	403	9	US-09-470-526-2	Sequence 2, Appl
23	103.5	2.8	525	9	US-09-470-526-9	Sequence 9, Appl
24	103.5	2.8	1114	10	US-09-782-874-2	Sequence 2, Appl
25	102.5	2.8	396	10	US-09-801-368-138	Sequence 138, App
26	102.5	2.8	833	9	US-10-014-436-3	Sequence 3, Appl
27	102.5	2.8	1242	9	US-09-925-299-911	Sequence 911, App
28	102.5	2.8	1242	10	US-09-925-299-911	Sequence 911, App
29	102.5	2.8	1938	9	US-10-014-436-2	Sequence 2, Appl
30	102	2.7	943	9	US-09-738-626-5199	Sequence 5199, Ap
31	102	2.7	951	9	US-09-924-097-15	Sequence 15, Appl
32	101.5	2.7	845	10	US-09-815-242-4972	Sequence 4972, Ap
33	101.5	2.7	845	10	US-09-815-242-10848	Sequence 10848, A
34	101	2.7	772	9	US-10-121-032-28	Sequence 28, Appl
35	101	2.7	772	9	US-10-093-037-28	Sequence 28, Appl
36	101	2.7	1706	10	US-09-864-761-46862	Sequence 46862, A
37	100	2.7	729	10	US-09-815-242-10132	Sequence 10132, A
38	100	2.7	4563	9	US-09-870-759-128	Sequence 128, Appl
39	99.5	2.7	722	10	US-09-765-272-84	Sequence 84, Appl
40	99	2.7	951	10	US-09-815-242-10465	Sequence 10465, A
41	99	2.7	2039	9	US-10-192-584-7	Sequence 7, Appl
42	98.5	2.6	637	9	US-09-997-868-4	Sequence 4, Appl
43	98.5	2.6	765	9	US-09-989-339-2	Sequence 2, Appl
44	98.5	2.6	944	10	US-09-815-242-11391	Sequence 11391, A
45	98	2.6	1213	10	US-09-815-242-5358	Sequence 5358, Ap

ALIGNMENTS

RESULT 1  
US-10-008-355-2  
; Sequence 2, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 712  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-2

Query Match	100.0%	Score	3719;	DB	9;	Length	712;
Best Local Similarity	100.0%;	Pred. No.	2e-285;				
Matches	712;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MQMKLSILLGALLLGASGVAKADKGMWLLNELNQENLDRMRELGFTLPLDLSLSFDPK	60				
Db	1	MQMKLSILLGALLLGASGVAKADKGMWLLNELNQENLDRMRELGFTLPLDLSLSFDPK	60				
Qy	61	SIANAVIFGGGCTGTTVSDQGLIFTNHHCGYGATQSQTVDHDLRDCFSVSRMTGCELP	120				
Db	61	SIANAVIFGGGCTGTTVSDQGLIFTNHHCGYGATQSQTVDHDLRDCFSVSRMTGCELP	120				
Qy	121	IPGLSVKYLKRTKVYTDKVEGOLKGITDMEKLRKAQEVQCQELAKKENADENQLCTIVEPF	180				
Db	121	IPGLSVKYLKRTKVYTDKVEGOLKGITDMEKLRKAQEVQCQELAKKENADENQLCTIVEPF	180				
Qy	181	YSNNEYFLIYDVFKDVRMVFAPPSVKGFGGTDNNMWPRTGDFSVFVYAGADNRPA	240				
Db	181	YSNNEYFLIYDVFKDVRMVFAPPSVKGFGGTDNNMWPRTGDFSVFVYAGADNRPA	240				

```
QY 241 EYSKDNKPKPVYFAAVSMOGYKADDYAMTIGPPGSTDRLYLTSMGWEDRIENENNRIEV 300
|||||
Db 241 EYSKDNKPKPVYFAAVSMOGYKADDYAMTIGPPGSTDRLYLTSMGWEDRIENENNRIEV 300
|||||
QY 301 RGIKQGIWEAMSADQATRIKYASKYAQSANYWKNSIGMNRGLARLDVIGRKRABERAF 360
|||||
Db 301 RGIKQGIWEAMSADQATRIKYASKYAQSANYWKNSIGMNRGLARLDVIGRKRABERAF 360
|||||
QY 361 DWTIRKNGKSAVYGDVLSLSEKAYKEGAKANREMTYLSSETLFGGTEVVVRAQAFANALATNP 420
|||||
Db 361 DWTIRKNGKSAVYGDVLSLSEKAYKEGAKANREMTYLSSETLFGGTEVVVRAQAFANALATNP 420
|||||
QY 421 DAHAGILKSLDDKYKDYLPKSLDRKVLPAMLDIVRRRIPADKLPDIFKNVIDKFKGDTKK 480
|||||
Db 421 DAHAGILKSLDDKYKDYLPKSLDRKVLPAMLDIVRRRIPADKLPDIFKNVIDKFKGDTKK 480
|||||
QY 481 YADFVFDKSVVPYSDKFHAMLKSMDEKFAKATEKDPAPVELSKSVIAAARAQAADAMANA 540
|||||
Db 481 YADFVFDKSVVPYSDKFHAMLKSMDEKFAKATEKDPAPVELSKSVIAAARAQAADAMANA 540
|||||
QY 541 YATEKGRKLFAGLRMPYGRALPSDANFTMRMSYGSIKGYEPQDGAWNYHTTGKGVLE 600
|||||
Db 541 YATEKGRKLFAGLRMPYGRALPSDANFTMRMSYGSIKGYEPQDGAWNYHTTGKGVLE 600
|||||
QY 601 KODPKSDEFAVQENILDLFRTKNYGRYAENGOLHIAFLSNNDITGNSGSPVFDKNGRLI 660
|||||
Db 601 KODPKSDEFAVQENILDLFRTKNYGRYAENGOLHIAFLSNNDITGNSGSPVFDKNGRLI 660
|||||
QY 661 GLAFDGNWEAMSGDIEFEPDLQRTISVDIRYVLFMDIKWQCPRLIOELKLI 712
|||||
Db 661 GLAFDGNWEAMSGDIEFEPDLQRTISVDIRYVLFMDIKWQCPRLIOELKLI 712
|||||

RESULT 2
US-10-008-355-8
; Sequence 8, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-8

Query Match 98.2%; Score 3652; DB 9; Length 699;
Best Local Similarity 100.0%; Pred. No. 3.8e-280;
Matches 699; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOWKLSILLGAALLGASGVAKADKGMWLLNLENOENLDRMRELGFTLPLDSLVSFDRP 60
|||||
Db 1 MOWKLSILLGAALLGASGVAKADKGMWLLNLENOENLDRMRELGFTLPLDSLVSFDRP 60
|||||
QY 61 STANAVVIFGGCGTGITVSDQGLIFTNHHCYGAIQSQSTVDHDLRDCGFSRTMGEELP 120
|||||
Db 61 STANAVVIFGGCGTGITVSDQGLIFTNHHCYGAIQSQSTVDHDLRDCGFSRTMGEELP 120
|||||
QY 121 IPGLSVKYLKRVKTDKVEGOLKGITDMEERLKAQEQVCEQLAKKENADENQLCIVEFP 180
|||||
Db 121 IPGLSVKYLKRVKTDKVEGOLKGITDMEERLKAQEQVCEQLAKKENADENQLCIVEFP 180
|||||
QY 181 YSNNEFLIVDYVFKDVRMVFAVPPSSVGKFGGTDNNMWRPHRTGDFSVFRVYAGADNRPA 240
|||||
```

```
Db 181 YSNNEFLIVDYVFKDVRMVFAVPPSSVGKFGGTDNNMWRPHRTGDFSVFRVYAGADNRPA 240
|||||
QY 241 EYSKDNKPKPVYFAAVSMOGYKADDYAMTIGPPGSTDRLYLTSMGWEDRIENENNRIEV 300
|||||
Db 241 EYSKDNKPKPVYFAAVSMOGYKADDYAMTIGPPGSTDRLYLTSMGWEDRIENENNRIEV 300
|||||
QY 301 RGIKQGIWEAMSADQATRIKYASKYAQSANYWKNSIGMNRGLARLDVIGRKRABERAF 360
|||||
Db 301 RGIKQGIWEAMSADQATRIKYASKYAQSANYWKNSIGMNRGLARLDVIGRKRABERAF 360
|||||
QY 361 DWTIRKNGKSAVYGDVLSLSEKAYKEGAKANREMTYLSSETLFGGTEVVVRAQAFANALATNP 420
|||||
Db 361 DWTIRKNGKSAVYGDVLSLSEKAYKEGAKANREMTYLSSETLFGGTEVVVRAQAFANALATNP 420
|||||
QY 421 DAHAGILKSLDDKYKDYLPKSLDRKVLPAMLDIVRRRIPADKLPDIFKNVIDKFKGDTKK 480
|||||
Db 421 DAHAGILKSLDDKYKDYLPKSLDRKVLPAMLDIVRRRIPADKLPDIFKNVIDKFKGDTKK 480
|||||
QY 481 YADFVFDKSVVPYSDKFHAMLKSMDEKFAKATEKDPAPVELSKSVIAAARAQAADAMANA 540
|||||
Db 481 YADFVFDKSVVPYSDKFHAMLKSMDEKFAKATEKDPAPVELSKSVIAAARAQAADAMANA 540
|||||
QY 541 YATEKGRKLFAGLRMPYGRALPSDANFTMRMSYGSIKGYEPQDGAWNYHTTGKGVLE 600
|||||
Db 541 YATEKGRKLFAGLRMPYGRALPSDANFTMRMSYGSIKGYEPQDGAWNYHTTGKGVLE 600
|||||
QY 601 KODPKSDEFAVQENILDLFRTKNYGRYAENGOLHIAFLSNNDITGNSGSPVFDKNGRLI 660
|||||
Db 601 KODPKSDEFAVQENILDLFRTKNYGRYAENGOLHIAFLSNNDITGNSGSPVFDKNGRLI 660
|||||
QY 661 GLAFDGNWEAMSGDIEFEPDLQRTISVDIRYVLFMDIKW 699
|||||
Db 661 GLAFDGNWEAMSGDIEFEPDLQRTISVDIRYVLFMDIKW 699
|||||

RESULT 3
US-10-008-355-9
; Sequence 9, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-9

Query Match 34.3%; Score 1276; DB 9; Length 720;
Best Local Similarity 38.9%; Pred. No. 2.2e-92;
Matches 283; Conservative 128; Mismatches 286; Indels 30; Gaps 11;

QY 6 KSTILGAALLGASGVAKADKGMWLLNLENOENLDRMRELGFTLPLDSLVSFDRP 65
|||||
Db 3 KRLLPLFAVLCLQIAHADEGMWLMQQLGRK-YAQMKERGLKMKREYDLYNPNGTSLKDA 61
|||||
QY 66 VVIFGGCGTGITVSDQGLIFTNHHCYGAIQSQSTVDHDLRDCGFSRTMGEELP 125
|||||
Db 62 VVIFGGCGTGITVSDQGLIFTNHHCYDMIOAHSTLEHNYLENGFWAMREADELPNKDTS 121
|||||
QY 126 VKYLRKIVKTDKVEGOLKGITDMEERLKAQEQVCEQLAKKE-----NADEQLCI-VEP 179
|||||
Db 122 VVIFDIEDVTDYVKKDLKAIKDPNSMDYLSPKYLOKLADKKRAGKNFSAPNGLSVLEIKA 181
|||||
```



[illegible]

Matches 245; Conservative 128; Mismatches 281; Indels 117; Gaps 25

QY 61 SIANVIFGGCGTGITVSDOGLFTNNHHCYGAIQSQSTVDHDLRDFGVSRMTGEELP 120

Query Match	21.7%	Score 807.5	DB 9	Length 732
Best Local Similarity	29.7%	Pred. No. 2.3e-55		
Matches	235	Conservative 119	Mismatches 289	Indels 147
Qy	8	ILLGAALLGASGVAKADKGMWLLNELNQENLDRMRELGTLPDLSLSYFDKPSIANAVV	67	
Db	3	IALLVAALVL-TCGATATADGEGQWQYQM-PSIADKLKSARGIDIPADKL--ADLTSYPMNAV	58	
Qy	68	IFGGCGTGITVSDOGLFTNHHCYCGALQSOSTVDHDLRGDFVSRVTMGELPIPGLSVK	127	
Db	59	VGLGYCTASFVSPGLVVTNHHCYAKAIQYNTKKHHNYLQSGFLATSDMKRE-PSAGPNER	155	
Qy	128	-YLRIKRVKTVTKVEGQLKGIT-DMEWRLKRAQECQELAKKENADENQOLCIVPEPFYSNNE	185	
Db	118	LYITE-AVTDVTSVTKLSQDLPKRYEELENIHSKALIKSCADDNYRCNVRSFHNGLE	175	
Qy	186	YFLIYVDFKDRMVFAPPSSVGRFGGDTDNMMPRHTGDFSVFRVYAGADNRPAEYSKD	241	
Db	176	YLLIKQLMIRDVRLVYAPPESVGGYGGIDINYEYPRHSGDFAFLRAYVGKDGKPAAPSED	235	
Qy	246	NKPYKPYFAAVSQGYKADDYAMTIGFPGSTDRY-LTS-----KGVEDRIENENNPR	297	
Db	236	NIPYTPKSYLVNADGVKAGDGVFAGYVPGTTRNYNTSELKFASDWLYPTQAKRY---Q	252	
Qy	298	IEVRIGIKQIWEAM-SADOATRICKYASKYAOSANYWKNSTGMNGLRLDVTGRKRAEE	356	
Db	293	LOIDTI-----EAMGQKDADIAIKYAGNWSMANRKKLNGLLAGFKATDIVGIIKOORE	346	
Qy	357	RAFADWIRKNGK-----SAVYGDVLSLSEKAY-----K	384	
Db	347	NDFLAWLTKPNLQNQLISEVLLEAQQLQTQNTYFTNAQGSTLTLLTAANNLYRLAKEK	406	
Qy	395	EGAKANREMTYLSSETLFGGTVEVRFQAFANALATNPDAHAGILKSLDDKYDYLPSLDRK	444	
Db	407	QKSDAEREIYQGERDL-----AMFSSR-----LKRIDSS-----SFDVK	439	
Qy	445	V-----LPAMLIDIVRRIRPADKLPIDFKNVTDKKFKGDTTKYADVFEDKSVVPYS DK	496	
Db	440	VDKTLWLQDLNAYLSQPNRVAALDNM-----LNLNDK-----NVSLAAKLDG	481	
Qy	497	PHAMLKSMDK-----EKFAKEK--KDPAVELSKSVIAAARAIAQAMANAYATEKGK	547	
Db	482	LVSLLTLTDQAORLAWMEADAKAFETSSDPPIRLAVALYDTNMAQEAKEILAGKLSTAR	541	
Qy	548	RLFFAGLREMYPGRALP--SDANTMRMSVGSIKGYEPQDGAWNYHTTCKGVLEKQDPK	605	
Db	542	PAYMAAVIDYKANNWPVYPDANGTLRISYGMWDVGYSRDALYKQPPTRLDGIVAKH-TG	600	
Qy	606	SDEFAVQENILDLFRTKNYGRY-----ANGOLHIAFTLSNND	642	
Db	601	VEPNYAPKKLLDITASVQRFGDHLVKSIVQDPGRGWICRLFSLCLKPEEFNSVPVNFUSSVD	660	
Qy	643	ITGGNSGSPVDFKNRGLTIGLAFDGNWEAMSGDIEFEFDLQRTISVDTRYLVLFMDIKWGQC	702	
Db	661	TGGNSGSPVNGKGLVGLNFDSTYEAITKDFEFTITRAVHVQDRIYLIWMWDEVSHA	720	
Qy	703	PRLIQELKLI	712	
Db	721	DNLIKELDLV	730	

RESULT 7  
US-10-008-355-3

```

RESULT 6
US-10-008-355-6
? Sequence 6, Application US/10008355
? Patent No. US20020164759A1
? GENERAL INFORMATION:
?
? APPLICANT: Travis, James
? APPLICANT: Potempa, Jan S
? APPLICANT: Banbula, Agnieszka
? TITLE OF INVENTION: Dipeptidylpeptidases
? FILE REFERENCE: 235, 004040101
? CURRENT APPLICATION NUMBER: US/10/008, 355
? CURRENT FILING DATE: 2001-11-08
? PRIOR APPLICATION NUMBER: US 607246, 827
? PRIOR FILING DATE: 2000-11-08
? NUMBER OF SEQ ID NOS: 26
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 6
? LENGTH: 732
? TYPE: PRP
? ORGANISM: Shewanella putrefaciens

```

; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-3

Query Match 7.4%; Score 274; DB 9; Length 52;  
Best Local Similarity 100.0%; Pred. No. 7.5e-15;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 644 TGGNSGSPVFDKNGRLGLAFDGNWEAMSGDIEPEPDLQRTISVDIRVLFPM 695  
1 TGGNSGSPVFDKNGRLGLAFDGNWEAMSGDIEPEPDLQRTISVDIRVLFPM 52

## RESULT 8

US-09-815-242-13382  
; Sequence 13382; Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13382  
; LENGTH: 883  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13382

Query Match 3.6%; Score 133; DB 10; Length 883;  
Best Local Similarity 20.4%; Pred. No. 0.059;  
Matches 104; Conservative 65; Mismatches 166; Indels 176; Gaps 24;  
Db 192 DVF-----KDVRYMFPAPSSVGKGGDTDNWPRHTGDFSVFRYAGADNRPAEYS 243  
26 DVFKPSGQKAKPYSIVPPNVYTKL-----HLG-----HA 57  
Qy 244 KDNKPKPVFAAASMOGYKADYAMITGPGSTDRYL-TSGWVEDRIENENNPRIVRG 302  
Db 58 WDT-TLQDIIIRKMQGFD-----TLWLPQMDHAGIATQAKVEERLGRGIESRYDLGR 110

Qy 303 ----IKQGIWKEAMSADQATRIKYA-SKYAOSANYWKSIGNRGLARLDVIGRKRAEER 357  
Db 111 ESFLTKVWENKD-----EVATTIKQWGMGLSVDSYRERFTLDEGLS-----KAVRK 158  
Qy 358 AFADWIRK-----NGKSAVYGD-----VLSSLEKAYK-----EGAKANREMT 394  
Db 159 VFVDLYKKGWYRGFTIINPDPAARTALSDIEVTHKDVGEAFYHMNYMLEGDSRALEVAT 218  
Qy 395 YLSETLFGGTEVVRFAOFANALATNPDAHAGILSKLDDKYKYDLPSS-----LDRKVLPM 449  
Db 219 TRPETMEGDV-----AIAVNP-----DPYKDLICKNVILPIANKLPIV 259  
Qy 450-LD-----IVRRIPADKLPDIFKNVDKFKGDTKKYADVFDPKSVVPYSOK 496  
Db 260 GDEHADPEFGTGVVVKITPAHDPNDFLVGQRHNLQVNMIMDDGTMLNLFVFEFSGM--DR 316  
Qy 497 FHAMLKSNDRKFAKALEKOPAVELSKSVIAAARA-----IQAD 535  
Db 317 FEAR-----RAVAKLEEIGALVKIEKRVHSVGHSERTGVVVEPRLSQTQWFMVKMDQLAKN 371  
Qy 536 AMANAYAIEKGKRLFFAGLREMPGRALPS-----DANFTWEXSYGSIKGYEPQDG 586  
Db 372 AIANQDTEDK-----VEFYPPRENDTFLQWNNVHDWVISRLWNGH-----QIP 416  
Qy 587 AWNYHTTGGVLEKQDPKSDFAVQENILD 617  
Db 417 AWYN--ADGENYVGEAEPEGGDGTQDEVDLD 445

## RESULT 9

US-09-815-242-13684  
; Sequence 13684; Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13684  
; LENGTH: 883  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13684

Query Match 3.4%; Score 127; DB 10; Length 883;  
Best Local Similarity 20.2%; Pred. No. 0.18; Mismatches 167; Indels 176; Gaps 24;  
Matches 103; Conservative 65; Mismatches 167; Indels 176; Gaps 24;

```
Qy 192 DVF-----KQVRFVFPAPSSVKGFGDQDNNMMPRHGTGDSFVRVYAGADNRPAEY 243
Db 26 DVFQSGQKAKPYIVIPPPNVTKL-----HLG-----HA 57
Qy 244 KDNKPYKPYFAAASMOGYKADDYAMTIGFGSGTDRL-TSWGVEDRIENENNPRIEVRG 302
Db 58 WDT-TLQDIIIRKRMQGF-----TLWLPQMDHAGATQAKVEERLGEIGITRYDLGR 110
Qy 303 ---IKQIGWEAMSADQATRIKYA-SKYAQSANYWKNISGNRGLARLDVIGRKRAEER 357
Db 111 ESFLTKVWENK-----EVATTIKEQWGMGLSVDSVSRERFTLDEGLS-----KAVRK 158
Qy 358 AFADWIRK-----NGKSAYGD-----VLSSLEKAYK-----ECAKANREMT 394
Db 159 VFVDLYKKGWIRGEFIINWDPAAARTALSDIEVIHKDVEGAFYHMNYMLEDGSRALVAT 218
Qy 395 YLSETLFGTGVVRFQAQFANALATNPDAHAGILSKDYKDYLP-----LDRKVLPA 449
Db 219 TRPETMFGDV-----AVAVNPE-----DPRYKDLIGKNVILPIANKLIPIV 259
Qy 450 LD-----IVRRRIIPADKLPDIEKNVIDKFKGDTKKYADFVFKSVVPYSDK 496
Db 260 GDERADPEFGTVGVKITPAHDNDELVCORHNLPOVNMDDGTWMLAFEGSM---DR 316
Qy 497 FAMLKSMDEKKEKAIEKDPAVELSKSVIAAARA-----IOAD 535
Db 317 FEAR-----KAVAKLEEIGALVKIEKRVHSGVHSERTGVVVEPRLSTQWFKMDOLAKN 371
Qy 536 AMANAYATEKGRLEFFAGLRMYPGRALPS-----DANFTMRMSYSGSIKGYEPDQG 586
Db 372 AIANQDTEK-----VEFYPPRFNDTFLQWMENVDHWVISRQLWVGH-----QIP 416
Qy 587 AWYNYHTTGKGVLEKQDKPSDEFAVOENILD 617
Db 417 AWYNY--ADGMYVGEAEPEGGTQDQEDVLD 445
```

## RESULT 10

```
US-10-008-355-18
; Sequence 18, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-18
```

```
Query Match 3.38; Score 122.5; DB 9; Length 26;
Best Local Similarity 96.28; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
Qy 24 ADKG-MWLLNENLNENLDRMRELFT 48
Db 1 ADKGWMLLNENLNENLDRMRELFT 26
```

## RESULT 11

```
US-09-925-301-1012
; Sequence 1012, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925.301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1012
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (153)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (229)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (433)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1012
```

```
Query Match 3.08; Score 113; DB 10; Length 708;
Best Local Similarity 19.28; Pred. No. 1.6;
Matches 101; Conservative 79; Mismatches 183; Indels 164; Gaps 28;
```

```
Qy 131 KIVKVTQVEGQLGITDEMERLKRKAQVCOEL-----AKKENADENQICIVEPPYSNNEY 186
Db 35 KVAKYKNKAPAEVO-ITAE-QLLREAKERELELLPPPPQKITDEEL-----NDY 83
Qy 187 FLIVYDFKDVRFVFPAPSSVKGFGDQDNNM---WPRHTGDSFVRVYAGADNRPAEY 242
Db 84 KLRKRKTEF-----NIRKNTVISNWKYAOW-----EES 114
Qy 243 SKDNKPYKPYFAAASMOGYKADDYAMTIGFGSGTDRLTSWGVEDRIENENNPRIEVRG 302
Db 115 LKEIQARSIIVERALDV-----DY-----RNITLWLKYAEMEMKNRQVXHAR- 156
Qy 303 IKQIGWEAMSADQATRIKYSKQAQSAQSNYKNSIGMNRGLARLDVIGRKRAEERAFADW 362
Db 157 ---NIWDRAI-----TTLPRVYNQFWKYTYMEMLG-----NVAGARQVFER---W 196
Qy 363 IRKNGKSAYGDLVLSLEKAYKEGAKANREMTYLSLSETLFGGTEV---VRFAQFANALA-- 417
Db 197 MEWOPEQAWHSYI-NFELRYKE---VDRARTIYERXVLVHPDVKNWIKYARFEKHAIFY 252
Qy 418 -----TNPDHAGI-LKSLDDKYKDYLPSLDRKVLPAMLDIVRRRIPAD 460
Db 253 AHARKVYERAVEFEFGDEHMDHLYVAPAKFEENQKEF---ERVRYIKYVALD---RISKQ 306
Qy 461 KLPIEFKN--VIDKKFGDGTKYADFVDFKSVVPSYDKFHAMLSMDK-EKFAKAIKDP 517
Db 307 DAQELFKNYTIFEKKF-GDRGIEDIIVSKRRFYEEVKANPHNYDAWFDYDLRVESDA 365
Qy 518 AVELSKSVIAAARAQADAMANAIAIEKGRLEFFAGLRMYPGRALPSDANFTMRMSYGS 577
Db 366 EAE-----AVREYERAIANVPPIQEKRH-----WKRYIY----- 395
Qy 578 IKGYEPQDGAWNYHTTGKGVLEKQDKPSDEFAVOENILDLEFRTKNY 624
Db 396 -----LWINVALVEE--LEAKDPERTRQVYQAS-LELIPHKFE 430
```

```
RESULT 12
US-10-192-584-6
```

; Sequence 6, Application US/10192584  
; Publication No. US20030027987A1  
; GENERAL INFORMATION:  
; APPLICANT: TOKUNAGA, Eiichi  
; SAKAGUCHI, Masashi  
; MATSUO, Kazuo  
; HAMADA, Fukusaburo  
; TOKIYOSHI, Sachio  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS  
; PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 624 Ninth Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10192,584  
; FILING DATE: 11-Jul-2002  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/077,098  
; FILING DATE: 19-May-1998  
; APPLICATION NUMBER: PCT/JP97/03222  
; FILING DATE: 12-SEP-1997  
; APPLICATION NUMBER: JP 27,148/1996  
; FILING DATE: 19-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KORNEAU, Anne M.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: TOKUNAGA-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2042 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-192-584-6  
Query Match 3.0%; Score 110; DB 9; Length 2042;  
Best Local Similarity 22.0%; Pred. No. 13;  
Matches 121; Conservative 78; Mismatches 188; Indels 164; Gaps 32;  
QY 184 NEYFLIVYDFKDVDMVFPAPSSVCKFGGDTNNMMPHRTGDFSVFR----- 230  
DB 2 NKVFRIKISVVKQEMIV-----VSELANNKDKTASQKNTHTAFFQPLFKTKTYLALLI 55  
QY 231 -VYAGADNRPAEYKDKNPYKPVYFAA5SMOQYKADD-----YANTI 271  
DB 56 NIALGASLFP---QLANAKWLEVYSSSVKSLTVSAQSNVNLNPSGAESVGTNSPQGVAI 112  
QY 272 GPPGSTDRYLT---SWGVEDRIENNNPRIEVRGIKQCIWKEAMSADQATRIKYASKYAQ 328  
DB 113 GYGATNDRSATCAIALGV--GVKNETLAKDSI-AIGYAKNESTAPSSVTICK-----Q 163  
QY 329 SANYWKNSI--GMNRLGLARLDVIGKRRAERAFADWIRKKNKSAVYGDVLSLEKAYREG 386  
DB 164 AINRFKSIIVMGLN-AYQLDPRGTSK-ETRGQSVVIGENAKSA--GN--QSVSLGQNSW 217  
QY 387 AKANREMYLTSETLGGTEVVRFAQAFANALATNPDAHAGILKSLDDKYKDYLP5LDRKVL 446  
DB 218 SKTN-----SISIGAGT---FAEGKSSIAIGTDKISG-----TKYNDKLPATAWNGT 261

QY 447 PAMLDIVRRRIPADKLPIDIFKNV-IDKKFKG---DTK-----KYADF-----VFD 487  
DB 262 GT-----VPKNSIMDIFSELYMGQTNGRDYDTTTRDPNKPFAFYKFSDFKGYKVT 313  
QY 488 KSVVP-YSDKFAHMLKSMDEKFAKTEKOPAVELSKSV--IAAARAIQADAMANAIAIE 544  
DB 314 PTASPTYAGKIGAI-----ALGSRTIAAGEMSTAVGSLAFALADRSTAM----- 357  
QY 545 KGKRLFFAGLRMPGPRALPSDA-----NFTMRMSYGSIX----- 579  
DB 358 -GLRSFVA--KDAVGCTAIGESRTFAKDSVAIGNKTEASNAGSWAYGKAKAYGAGAIA 414  
QY 580 -GYEPQDGAWNYHTTGKGVLE-----KQDPKSDEFAVOENTILDLFRTKNRYGAENG 631  
DB 415 ICTEVAAGAKFNHQTNLLQDNNAATLKNADKSDS-TKTGNATVTF-TQSFEDNMLTNG 472  
QY 632 QLHIAFLSNND 642  
DB 473 ---LPLVSENE 480  
RESULT 13  
US-10-027-806-62  
; Sequence 62, Application US/10027806  
; Patent No. US20020160476A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Schleper, Robert A.  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCCORP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,806  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 62  
; LENGTH: 845  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-027-806-62  
Query Match 2.9%; Score 109.5; DB 9; Length 845;  
Best Local Similarity 19.1%; Pred. No. 4;  
Matches 112; Conservative 99; Mismatches 209; Indels 165; Gaps 28;  
QY 20 GVAKADKGMWLLNE-----LNQENLDRMRELGTLPDLSLYSF-----DKP5IANAVVI 68  
DB 301 GVAQSDIPLYMMRDSATLRHGVHLDLYRT--FSNRSFQYAFAAKYTDYSLNSVSKAMLG 358  
QY 69 FGGCTGTIVSDQGLIFTNHHCYGA--IQSQSTVDHDLRDGFVSRVTMGBELPIPLGLS- 125  
DB 359 ECKVDYGVSLGDLTLTYQTANYCYHDARLTLELSTFGNEILMDLLVVT5RIARMPIDDSR 418  
QY 126 ---VKYLRKIVKVTDKVEGQLGITDEMRUKRAQVCOELAKKENADENOLCIVEP--- 179  
DB 419 MCVSOWIRSLIYYEHRQNALIPRDELE--KRSQOVSDNAVICKKFRGGL-VVEPEG 475  
QY 180 -----FYSNNEYFLIVYDV-FKDYRMVFPAPSSVCKFGGDTNNMMPHRTG----- 224  
DB 476 IHFDVTVMDFASLYPSIIKVRNLSYETVRCVH--PECRKNTIPDTHWVCTKNGLTSMI 533  
QY 225 -----DF-----SVFRVYAGADNRPAEYKDKNPYKPVYFAA5SMOQYK-----AD 265  
DB 534 ICSLRDLRVNYKSLSKSQSITEQROQYTVISQALKVVLN5ASYGVMAEIPFLYFLPAA 593  
QY 266 DYAMTIGFPGSTDRYLT5WGVEDRIENNNPRIEVRGIKQCIWKEAMSADQATRIKYASK 325  
DB 594 EATTAVG-----RIIIMQTI5-----HCEQMGVK-----VLYGD- 622  
QY 326 YAQSAANYKNSIGMNRGLARLDVIGKRRAERAFADWIRKKNKSAVYGDVLSLEKAYKE 385

Db 623 -----TDSLFIKNPEERQIHDIVHAKKE--HG-VELEVDEKEYRY 659  
Qy 386 GAKANREMTYLSSETLFGGTEVVRFAQANALATNPDAHAGILKSDDKYKDYLPISLDRKV 445  
Db 660 VVLSNRKKNYFGVTKSGKVDV-----KGLTGK-KSHTPPFIKEL 697  
Qy 446 LPAMLDIVRRRIPADKLPDIFKNV---IDKFKGDTKKYADVFDKSVVPYSD-KFHAML 501  
Db 698 FYSLLDI-----LSAVQTEDEFESAKLKISKAIAASGKR-----LEERGVLADLAFNVI 748  
Qy 502 KSMDEKFAKATEKDPAVELSKSVIAAARAIQADAMANAYALEKG 546  
Db 749 SKAPSE-VYKTPVQH-----IRAARLLE-----NAREVKKG 778

## RESULT 14

US-10-034-623-62  
; Sequence 62, Application US/10034623  
; Publication No. US20020198365A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCCP.002A  
; CURRENT APPLICATION NUMBER: US/10/034,623  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 60/102,294  
; PRIOR FILING DATE: 1998-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 62  
; LENGTH: 845  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-034-623-62

Query Match 2.9%; Score 109.5; DB 9; Length 845;  
Best Local Similarity 19.1%; Pred. No. 4;  
Matches 112; Conservative 99; Mismatches 209; Indels 165; Gaps 28;

Qy 20 GVAKADKGMWLLNE---LNOENLDRMRELGTLPDLSLYSF-----DKPISANAVVI 68  
Db 301 GVADSDIPLYMMRDSATLRHGVLHDLRYT--FSNRSFQIYAFAAKYDYSLNSVSKAMLG 358  
Qy 69 FGGGCTGTVSDQGLIFTNHHGCGYA--IQSQSTVDHDLRDLGDFVSRMTGEBELPIPGLS- 125  
Db 359 EGKVDYGVSLGDLTLTYQTANYCYHDARLTLELSTFGNEILMDLLVVTSGRIARMPIDDKSR 418  
Qy 126 ---VKYLRKIVKVDKVEGOLKGIIDEMERLKAQEVCOELAKKENADENQICIVEP--- 179  
Db 419 MGVSWIRSLIYEHQRNALIPRDELE--KRSQVSNDAVIKDKFRGGL-VVEPEEG 475  
Qy 180 -----FYSNNEYFLIVYDV-FKQVRYMFAPSSVSGKFGGDDNNMWRPHGTG----- 224  
Db 476 IHFDVTVMDFASLPSIITKVRNLSYETVRCVH--PECKRNTIPDTHHWCTKNGLTSMI 533  
Qy 225 -----DF-----SVFRVYAGADNRPAYSKDNPKYPVFAAVSMQGYK-----AD 265  
Db 534 IGSRLDRLVNYKSLKSQSITTEQROQYTVLSQALKVVLNASTYGVMAEIPFLYFLPAA 593  
Qy 266 DYAMTIGPGSTDRLTYSWGVEDRIENENNPRIEVRGKIQGIWKEMASDAQTRIKYASK 325  
Db 594 EATTAVG-----RYIIMQTIS-----HCEQMGYK-----VLYGD- 622  
Qy 326 YAQSANYNKNSIGMNRGLARLDVIGRKRAERAFADWIRKNGKSAVYGDVLSLSEKAYKE 385  
Db 623 -----TDSLFIKNPEERQIHDIVHAKKE--HG-VELEVDEKEYRY 659  
Qy 386 GAKANREMTYLSSETLFGGTEVVRFAQANALATNPDAHAGILKSDDKYKDYLPISLDRKV 445  
Db 660 VVLSNRKKNYFGVTKSGKVDV-----KGLTGK-KSHTPPFIKEL 697  
Qy 446 LPAMLDIVRRRIPADKLPDIFKNV---IDKFKGDTKKYADVFDKSVVPYSD-KFHAML 501  
Db 698 FYSLLDI-----LSAVQTEDEFESAKLKISKAIAASGKR-----LEERGVLADLAFNVI 748

Db 660 VVLSNRKKNYFGVTKSGKVDV-----KGLTGK-KSHTPPFIKEL 697  
Qy 446 LPAMLDIVRRRIPADKLPDIFKNV---IDKFKGDTKKYADVFDKSVVPYSD-KFHAML 501  
Db 698 FYSLLDI-----LSAVQTEDEFESAKLKISKAIAASGKR-----LEERGVLADLAFNVI 748  
Qy 502 KSMDEKFAKATEKDPAVELSKSVIAAARAIQADAMANAYALEKG 546  
Db 749 SKAPSE-VYKTPVQH-----IRAARLLE-----NAREVKKG 778

## RESULT 15

US-10-027-801-62  
; Sequence 62, Application US/10027801  
; Publication No. US20030054364A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCCP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,801  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 62  
; LENGTH: 845  
; TYPE: PRT  
; ORGANISM: Cenarchaeum symbiosum  
US-10-027-801-62

Query Match 2.9%; Score 109.5; DB 9; Length 845;  
Best Local Similarity 19.1%; Pred. No. 4;  
Matches 112; Conservative 99; Mismatches 209; Indels 165; Gaps 28;

Qy 20 GVAKADKGMWLLNE---LNOENLDRMRELGTLPDLSLYSF-----DKPISANAVVI 68  
Db 301 GVADSDIPLYMMRDSATLRHGVLHDLRYT--FSNRSFQIYAFAAKYDYSLNSVSKAMLG 358  
Qy 69 FGGGCTGTVSDQGLIFTNHHGCGYA--IQSQSTVDHDLRDLGDFVSRMTGEBELPIPGLS- 125  
Db 359 EGKVDYGVSLGDLTLTYQTANYCYHDARLTLELSTFGNEILMDLLVVTSGRIARMPIDDKSR 418  
Qy 126 ---VKYLRKIVKVDKVEGOLKGIIDEMERLKAQEVCOELAKKENADENQICIVEP--- 179  
Db 419 MGVSWIRSLIYEHQRNALIPRDELE--KRSQVSNDAVIKDKFRGGL-VVEPEEG 475  
Qy 180 -----FYSNNEYFLIVYDV-FKQVRYMFAPSSVSGKFGGDDNNMWRPHGTG----- 224  
Db 476 IHFDVTVMDFASLPSIITKVRNLSYETVRCVH--PECKRNTIPDTHHWCTKNGLTSMI 533  
Qy 225 -----DF-----SVFRVYAGADNRPAYSKDNPKYPVFAAVSMQGYK-----AD 265  
Db 534 IGSRLDRLVNYKSLKSQSITTEQROQYTVLSQALKVVLNASTYGVMAEIPFLYFLPAA 593  
Qy 266 DYAMTIGPGSTDRLTYSWGVEDRIENENNPRIEVRGKIQGIWKEMASDAQTRIKYASK 325  
Db 594 EATTAVG-----RYIIMQTIS-----HCEQMGYK-----VLYGD- 622  
Qy 326 YAQSANYNKNSIGMNRGLARLDVIGRKRAERAFADWIRKNGKSAVYGDVLSLSEKAYKE 385  
Db 623 -----TDSLFIKNPEERQIHDIVHAKKE--HG-VELEVDEKEYRY 659  
Qy 386 GAKANREMTYLSSETLFGGTEVVRFAQANALATNPDAHAGILKSDDKYKDYLPISLDRKV 445  
Db 660 VVLSNRKKNYFGVTKSGKVDV-----KGLTGK-KSHTPPFIKEL 697  
Qy 446 LPAMLDIVRRRIPADKLPDIFKNV---IDKFKGDTKKYADVFDKSVVPYSD-KFHAML 501  
Db 698 FYSLLDI-----LSAVQTEDEFESAKLKISKAIAASGKR-----LEERGVLADLAFNVI 748







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2003, 15:31:51 ; Search time 44.9435 Seconds  
(without alignments)  
1522.972 Million cell updates/sec.

Title: US-10-008-355-2  
Perfect score: 3719  
Sequence: 1 MQMKLSILGALLGASG.....LFMIDKWCOPRIQELKLI 712  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	907.5	24.4	716	2 G82627	hypothetical prote
2	133.5	3.6	1104	2 A60999	alpha-amyrase (EC
3	133	3.6	883	2 D97933	valine-tRNA ligase
4	129	3.5	1104	1 A36866	microbial collagen
5	125	3.4	1440	2 T33813	hypothetical prote
6	124.5	3.3	735	2 F98228	1,4-alpha-glucan b
7	124.5	3.3	735	2 AH3057	glycogen branching
8	123.5	3.3	776	2 S44784	C30C11.4 protein -
9	123	3.3	1289	2 F72308	hypothetical prote
10	122.5	3.3	1571	2 AC1647	Lactobacillus phag
11	122	3.3	721	2 AD1617	penicillin-binding
12	122	3.3	743	2 D82883	DNA helicase II UU
13	122	3.3	883	2 C95066	valyl-tRNA synthet
14	120	3.2	666	1 D69103	DNA helicase (EC 3
15	118	3.2	698	2 C82332	translation elonga
16	116	3.1	952	2 T36664	probable DNA topoi
17	115.5	3.1	1154	2 F71856	hypothetical prote
18	114.5	3.1	561	2 G83913	hypothetical prote
19	114.5	3.1	1198	2 T42223	ladder protein - C
20	114.5	3.1	1306	2 S26224	aggregation protei
21	114.5	3.1	1530	2 H83865	glutamate synthase
22	114.5	3.1	1758	2 T34393	hypothetical prote
23	114	3.1	620	2 S19660	dnak-type molecula
24	114	3.1	620	2 S73236	transferrin-bindin
25	114	3.1	711	2 T06660	internal virion pr
26	114	3.1	747	1 HIBPC7	DNA topoisomerase
27	113.5	3.1	765	2 H81805	pyruvate formate-1
28	113	3.0	759	2 AE1686	ribonucleoside-dip
29	113	3.0	1047	2 F81728	

ALIGNMENTS

RESULT 1  
G82627

hypothetical protein XFI887 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C;Accession: G82627  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: G82627  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-716 <SIM>  
A;Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AAF84693.1; GSPDB:GN  
A;Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
chado, M.A.; Madeira, A.M.F.; Marino, C.L.; Marques, M.V.; Martins  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XFI887

Query Match 24.4%; Score 907.5; DB 2; Length 716;  
Best Local Similarity 31.4%; Pred. No. 4.8e-51;  
Matches 239; Conservative 135; Mismatches 291; Indels 95; Gaps 21;

QY	1	MQMKLSILGALLGASGVAKADKQWLLNELNQNELDRMRELGTPLPDSLSYSDKP	60
DB	1	MRFNLLSLVLATLITVDS--THAGEMWVPOQL-PEIAGPLKQAGLQSLPEQLSNLTGD	57
QY	61	SIANAVIFGGCTGTVSDQGLFTNHHCYGAIQSOSIVDHYLDGDFVSRMTGELP	120
DB	58	PMGAVVSL--GNCTASIVSPEGLVITNHHCAYGAIQSNLTPKKNLKEGFFNALQADEVS	115
QY	121	I-PGLSVKYLKRVKVDKVEGQKLGITDEMERLKAQEV--QELAKKENADENQLCIV	177
DB	116	AGPNARIYVLEQITDVTAAQAKAALAAAGNDPFKRTTALFTFSKQEIACEE-EGYRCQF	174
QY	178	EPFYSNNEYFLIVVDKPRVMVFAPPSSVCGKFGGDTNNMWRHTGDFSVFRVYAGADN	237

2',3'-cyclic-nucle  
topoisomerase IV c  
type I restriction  
hypothetical prote  
esterase - Caenorh  
hypothetical prote  
DNA-directed DNA p  
aconitate hydratase  
hypothetical prote  
hypothetical prote  
Leu/Val-bindin  
fibronectin-bindin  
probable phage tai  
protein L precursor  
isoleucine-tRNA li  
restriction modifi

```
Db 175 FSPAGNTYRVFNKLEIKDVLRYVAFQSGVGKGGDDVNNMWRHTGDFSFYRAYVGKDG 234
QY 238 RPAEYKDNKPKYVPVFAAVSMQYKADYAMTIGPFGSDRYLTSGWGEDRIENENNR 297
Db 235 KPASFSENIPYRPHKLFSDQPLGDDGVVAGYGRNRYALV-----AEFENTAH 288
QY 298 IEVRGKIQ-----GIWKEAMSADQATRIKYAKSQAQSYWNSIGMNRGLARLDVIGR 351
Db 289 WTPYVIGQHFKNLIALTEAASKNPDIQVYASTAGLNNTSKNFQDGLDFRRINAIG 348
QY 352 KRAERAFADWIRKNG-----KSAYVGDVLSLEKAYKEGAKANREMTYLSSETLFGGTEVV 407
Db 349 KQSEETAVLAWLKQGGIRGHEALAAHOTLVLDLTEQY-----KANQDRDFVLGQ-FNGSGVI 403
QY 408 RFAQFANALA-----TNPDAA-----AGILKSLDDKYLDKYLPSLDRKV-----445
Db 404 GVAVNLRYLAERTKSAQREAGYQERDLPTIEGNLKQME---RRYLPENDRQMQVWLT 460
QY 446 ----LPAMLDIVRRRTPA-----DKLPDIFKNVVDKFKGDTKKYADFVDFKSVVPYSD 495
Db 461 EYNKLP-----VKQVAAIDVWLGDGIPATLKRIGDPTKLSSSEB-----500
QY 496 KFHAMLKSMDEKFAKAEKDPAVELSKSVIAAARATQADAMANAIAIEKGRKLFPAGLR 555
Db 501 ----LKNFNADRAAFESSODPAIRYAVAIIMPALLEIERONKIRTGELLKARPIYLOALA 555
QY 556 EYYP--GRALPDSANFTMRMSYSGIKYEPQDGAWYNYHTTGRGVLEKQDPKDSDEFVQVE 613
Db 556 DYNSHGKFFVYPDANSSLRITFGHVGYSPKDGVEYTFPTTLCQVMAK-NTGVEPEDSPK 614
QY 614 NIDLPFTKNYGRYAEN--COLHIAFLSNNDITGNGSGSPVDFKNGRGLIGLAFDGNWEAM 671
Db 615 SLINAIKAKSYANLADQIRGTVFVNFUSLDITGNGSGSPVLDAGHKLVLGAFDGNWESV 674
QY 672 SGDEPEPDLQRTISVDIRVFLFMIDKMGOCPLRIQELK 711
Db 675 SSNWVDPVMTRIANDSRVQVIMTEVAPAPHLKELNL 714

RESULT 2
A60999
alpha-amylase (EC 3.2.1.1) precursor - Micrococcus sp. (strain 207)
C:Species: Micrococcus sp.
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C:Accession: A60999
R:Kimura, T.; Horikoshi, K.
FEMS Microbiol. Lett. 71, 35-42, 1990
A:Title: The nucleotide sequence of an alpha-amylase gene from an alkalopyschrotrophic M
A:Reference number: A60999
A:Accession: A60999
A:Molecule type: DNA
A:Residues: 1-1104 <KIM>
A:Cross-references: GB:X55799; NID:g296762; PIDN:CAA39321.1; PID:g296763
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-1104/Product: alpha-amylase #status predicted <MAT>

Query Match 3.6%; Score 133.5; DB 2: Length 1104;
Best Local Similarity 20.8%; Pred. No. 1.9;
Matches 138; Conservative 94; Mismatches 259; Indels 173; Gaps 35;

QY 141 GOLKGIITDNERLKAQEVQELAKKENADENQLCIVEPFSYNNNEFLIVYDFKQVRMV 200
Db 384 GDFKGIITDKLDYL-----DELGVNTIWI-SPVVENIKY-----DVRYY 420
QY 201 FAPPSSVKGFGDQDNM--MWRPHGTGDFSVFR-VYAGADNRPAEYSKDNKPKYKPVFAAV 257
Db 421 ETSEPPYGYGHWANNFEGELNP-HFGTMEEFHDLIDGAHNRNMKIMVD-----VVV 470
QY 258 SMQGY--KADDYAMT--IGFPGSGTDRLYS-----NGVEDRIENEN 294
```

```
Db 471 NHTGYGLKEIDGSGVTNPPAGYPSDADRAREFSDLLROGADVGTDEVVGLAGLDFDITFD 530
QY 295 NPRIEVRGIRKQGTWKEAMSADQATRIKY-----ASKYAQSAQSYWNSIGMNRGLARLDVIG 350
Db 531 NVRKQIID-WQTDWIEKATTENGNTIDYFRVDTVKHVEDATWQFKNALTEKPEFKMIG 589
QY 351 KRAREAFADWIRKNGKSAYV---GDVLSLEKAYKEGAK-----ANREMTYLS 398
Db 590 EA-----WGAKVDNTLGLYLETGTMTDLSLDFGFKETARSEFVNGSLEAANASLTARNA 640
QY 399 TLEGGTEVVRFB-----AQFANALATN-----PDAGAGILKSLDDKYY 435
Db 641 KLNTATLFGFLGSHDEEGLHSLAGDKGLQVAAATLQATAKQGPVIYYG--BELGOTGA 698
QY 436 DYLPSLDRKVLPAMLDIVRRRIIPADKLPIFKNVVDKFKGDTKKYADFVDFKSVVPYSD 495
Db 699 NNPQYDNR-----YDFAWDOVEGNETLAHYTKILNFR-EGYSKVFVK--GERTLVGSD 750
QY 496 KFHAMLKSMDEKFAKAEKDPAVELSKSV---IAAARATQADAMANA-YAIEKGRKLF 551
Db 751 KDOFLFSRDYQDQKVVYVGLNVAEE-SKAVTLTVDSADAVVTDAYSCTEYTAGK----805
QY 552 AGLREMYVGRALPDSANFTMRMSYSGIKG-----YEPODGAWYN 590
Db 806 --VNLTPLGRADGTVLLT--VEGNGITGVAKONGEVVVELVPENNIRIHYKREDNVYKN 861
QY 591 YHTTGKGVLEKQDPKDSDEFVQVQENILDLFRTKNRYAE-----NGQLHIAFLSNNDITGG 646
Db 862 Y--GAWLNVDVASPSANWVPVGTATMFE--KTDSYGAVIDVPLADGAKNIGFLV-MDITAG 915
QY 647 NSSPVEDKNGRLIGLAFDGNWENAMSGD--IEFEP-DLQRTISVDIRVFLMID-----697
Db 916 DAGKDGDKGFTTSSPOANEIWKQSGDKVYTYEVPVLDPAN-TVRIHYTREAVDYDDFGI 974
QY 698 -KWG 700
Db 975 WNWG 978

RESULT 3
D97933
valine-tRNA ligase (EC 6.1.1.9) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 03-Jun-2002
C:Accession: D97933
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D97933
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-883 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99296.1; PID:g15458063; GSPDB:GN00174
C:Genetics:
A:Gene: valS
C:Superfamily: valine-tRNA ligase
C:Keywords: ligase

Query Match 3.6%; Score 133; DB 2: Length 883;
Best Local Similarity 20.4%; Pred. No. 1.4;
Matches 104; Conservative 65; Mismatches 166; Indels 176; Gaps 24;

QY 192 DVF-----KDRMYFAPPSSVKGFGGDTDNWWRPHHTGDFSVFRVYAGADNRPAEYS 243
Db 26 DVFKPSGQKAKPYSVIIPPNVTGKL-----HLG-----HA 57
QY 244 KDNKPKYKPVFAAVSMQYKADYAMTIGFPGSGTDRLYL-TSWGVEDRIENENNPRIEVRG 302
```

Db 58 WDT-TLQDIIHKKRMQGF-----TLWLPQMDHAGIATQAKVEERLGRGEGISRYDLGR 110  
QY 303 ----IKGIWEKMSADQATRIKYA-SKYAOSANYWKNSIGNRGLARLDVIGRKRAERAFADW 357  
Db 111 ESFLKVVWEKD-----EYATTIKEQWKGMLSDVYSRERFTLDEGLS-----KAVRK 158  
QY 358 APADWIRK-----NGKSAYGD-----VLSSLEKAYK-----EGAKANREMT 394  
Db 159 VFDVLYKKGWIYRGFFIINWDPAARTALSDIEVIHKDVEGAFYHMYNMLEDCSRALEVAT 218  
QY 395 YLSETLFGTEVVRFAQANALATNPDAHAGILKSLDDKDYKDYLPSS-----LDRKVLPLAM 449  
Db 219 TRPETMFGDV-----ALAVNPE-----DPRKDLICKNVLPIANKLPIV 259  
QY 450 LD-----IVRRRIPADRLPDIFKNVIDKFKGDKFYADFVKFSVVPYSDK 496  
Db 260 GDEHADPEFGTVGVKITPAHDNDFNLVGRHNLPOVINMDDGTNMLVFESGM--DR 316  
QY 497 PHAMUKSMDEKFAKAIKDPVELSKSVIAAARA-----IQAD 535  
Db 317 FEAR-----KAVVAKLEEIGALVKIEKRVHSGVHSERTGVVVEPRLSQWFVKMDQLAKN 371  
QY 536 AANAYATIEKGRKFLFAGLREMYPGRALPS-----DANFTMRMSYGSIKGYEPQDG 586  
Db 372 AIANDQTEK-----VEYPPRFNDTFLQWENVHDWVISQLMWGH-----QIP 416  
QY 587 AMYNYHTTGKGVLEKQDKPSDEFAYQENILD 617  
Db 417 AMYN-ADGEMYVGEAPEGDGWTODEVDLD 445

RESULT 4  
A36866  
microbial collagenase (EC 3.4.24.3) precursor - Clostridium perfringens  
C:Species: Clostridium perfringens  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: A36866  
R: Matsushita, O.; Yoshihara, K.; Katayama, S.I.; Minami, J.; Okabe, A.  
J. Bacteriol. 176, 149-156, 1994  
A: Title: Purification and characterization of a Clostridium perfringens 120-kilodalton  
A: Residues: 1-1104 <MAT>  
A: Cross-references: GB: D13791; NID: g440850; PIDN: BAA02941.1; PID: g440851  
A: Gene: colA  
C: Superfamily: microbial collagenase  
C: Keywords: hydrolase; metalloproteinase

Query Match 3.5%; Score 129; DB 1; Length 1104;  
Best Local Similarity 18.2%; Pred. No. 3.6;  
Matches 142; Conservative 106; Mismatches 276; Indels 258; Gaps 36;  
QY 29 WLLNELNENL-----DRMRLEGLTLPDLSYFDPKPSIANAVIFGGCT---GITVSD 80  
Db 419 MASKEVKAQFMRVQNDKALEEGNPDILTIVVINSPEEYKLNRIINGFSTDNNGIYIEN 478  
QY 81 OGLIETNHHCYGAIOS-QSTVDHD---YLRDGFYSRTINGEELPPLGL---SVKYLKIV 133  
Db 479 IGTFYERTPEESYITUEELPRHEFTHYLQGRYV-----VPGMWGQGEFYQEGVL 529  
QY 134 KVTDRVEGOLKITDEMELRKAQVCEQLAKKENADENQOLCIVEPFYSNNEYFLIVDV 193  
Db 530 TWYEGSTAEFFAGSTRTDGIKRPSVTOGLAYDRNRRNSLYVLHAKYGSWDFY-----583  
QY 194 FKDRVRWFAPSSVGKFGGTDNNMWRPHRTGDFSVFRVYAGADNRPAYSKDNKPKPY 253  
Db 584 -----NYGFALSNNYNNMGMF-----NKWTNYIKNN-----611  
QY 254 FAASVWQYK-----ADDYAMTIGFPGSTDRYLTLSWGVEDRIENENNPRIE---VRGIK 304

Db 612 ----DVSGYKDYIASMSDYGLNDKYQDYMDSLL-----NNIDNLDVPLVSDEYVNG-- 659  
QY 305 QGIWEKMSADQATR-IKYASKYAO-SANYWKNSIGMNRGLARLDVIGRKRAERAFADW 362  
Db 660 ----HEARDINEITNDIKEVSNIKDLSNVKESQSFPTTYDMRGTYVGGRSQGEEN---DW 712  
QY 363 IRKNGKSAYVGDVLSLEK-----AYKEGAKANREMTYLSSETLFGTEVVRFAQF 412  
Db 713 KDMNSK---LNDMLKELSKSWNGYKTVTAYFVNHKVDENGNYVYDVVPHGMNT-----763  
QY 413 ANALATNPDAHA-----GILKSLDDKDYKDYLPSLDRKVLPLAMLDIVRRRIPADKLPDIFK 467  
Db 764 ----DTNTDVBHVNKPKAVIKS-----DSSV-----IVEEIN-----ED 794  
QY 468 NVIDKFKFGDTKKY-----ADEVFDKSVVPYSDKFAHMLKSM-----K 506  
Db 795 GFESKDEDEGEIKAYEWDGEGKSNEAKATHYKN-----TGEYEVKLVITDNNGGINT 849  
QY 507 EKFAKAIKDP-----AVELSKSVIAAARAIAQADAMANAAYAIEKGR-----548  
Db 850 SKKIKVVEDKPEVINESEPNNDFEKANQIAKSNMLVKGTLSSEYSDKYFYDVAKKGNV 909  
QY 549 -----LFFAGLREMY-----PGRALPSDANFTMRMSYGSIKGYEPQDG 586  
Db 910 KTTLNLSVGTITWTLYKEGDLNNVLYATGNDGTTELKGEKTELEPGYYLSVYTDNQSG 969  
QY 587 AMYNYHTTGKGVLEKQDKPSDEFAYQENILDLFRTKNYGRYAENGOLHIAFLSNNDITGG 646  
Db 970 A---YTVNVKGNLKNEVKETEKAKE-----VEN-----NND-----999  
QY 647 NSGSPVFDK-----NGRLIG-LAFDGNWMEASGDIEFEPDLQRTI---SVDIRYVLFM 695  
Db 1000 -----FDKAMKVDSNSKIVGTLSNDDLDKIYSIDIQNPSDLNIVVENLDNKMNNLLYS 1053  
QY 696 ID 697  
Db 1054 AD 1055

RESULT 5  
T33813  
hypothetical protein VC5.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
A: Accession: T33813  
R: Tin-Wollam, A.M.; Wohldmann, P.  
submitted to the EMBL Data Library, November 1998  
A: Description: The sequence of C. elegans cosmids VC5.  
A: Reference number: 221417  
A: Accession: T33813  
A: Status: preliminary; translated from GB/EMBL/DDBJ  
A: Molecule type: DNA  
A: Residues: 1-1440 <TIN>  
A: Cross-references: EMBL: AF106581; PIDN: AAC78208.1; GSPDB: GN000023; CESP: VC5.3  
A: Experimental source: strain Bristol N2; clone VC5  
C: Genetics:  
A: Gene: CESP: VC5.3  
A: Map position: 5  
A: Introns: 15/3; 27/2; 258/3; 382/3; 506/3; 630/3; 754/3; 879/3; 1005/3; 1410/3

Query Match 3.4%; Score 125; DB 2; Length 1440;  
Best Local Similarity 20.1%; Pred. No. 9.9;  
Matches 108; Conservative 72; Mismatches 162; Indels 194; Gaps 30;  
QY 132 IVK-VTDKVEGOLKITDEMELRKAQVCEQLAKKENADENQOLCIVEPFYSNNEYFLIV 190  
Db 237 LVKGVVDRODGEVKVIAEKM-----LSVCGEYKESTRKR-----271  
QY 191 YDVFEDKRVWFAPPSSVGKFGGTDNNMWRPHRTGDFSVFRVYAGADNRPAYSKDNKPKY 250  
Db 272 ----REIEAFAQ-----DFVKWMTPEQLGDITALKA-AGKDEVOAKVKE-----311  
QY 251 PVYFAAVSMQYKADYAMTIGFPGSTDRYLTLSWGVEDRIENENNPRIEVRGKIQGIWKE 310



QY	248	PKPVTFAAVSMQG----	YKADYA-----	MTIGFPGSTDYRLYSWGVEDRIENEN	294	
		:		:   :	:	
Db	196	AYK--FEILGANGELPLKADPYARRGELRPKNASVTAPELTQK-----	EDQAHREH	246		
QY	295	NPRIEYR-----	GLKQGIWKEA-----	MSADO--ATRIKYASK-----	325	
		:	:	:	:	
Db	247	WAQVDDORROPISIEYFHAGSWSORREDTGLTSDWEALAQLIPYCTDMGTFHTHIEFLPITBHP	306			
QY	326	-----	YAQSANTYKWSIGNMRGLARLDIGRKRAERAFADWI-----	363		
		:	:	:	:	
Db	307	YDPSWCYQTTGLYAPTARE-----	GDPEGFARF-VNGAHKVGIGVLLDWVPAAHPTDEHG	360		
QY	364	-----	RKNGKSAVYG-----	DVLSL-----	EKAYKEGAK--AN	390
		:	:	:	:	
Db	361	LRWFDGTALYEHADPRQGFPHDPWNTAIYNFGRIEYMSYLINNALYWAEBKFLDGLGRVDAV	420			
QY	391	REMYTILSETL-----	FGG-----	TEVYRFAQAFANALATNPDAHAGILKSLDDKYKDYL	438	
		:	:	:	:	
Db	421	ASMLYLDYSRKECEWIPNEYGGRENLESVRFLOKNNLSLYG--	THPGVMTAEESTS--W	476		
QY	439	PSLDRKVLPAMLDIVRRRIIPADKLPIFNKVIDKKFKGDKTKYADVFEDKSVVPYSDFKH	498			
		:	:	:	:	
Db	477	PKYSQPVHEGGLG-----	FGFKNNMGFMHDTLSY-----	FSRE--PVHRKPH	516	
QY	499	AM-----	LKSMDEKEKAKATEKDPVELSKSVIAAARAIQADAMANATAI-----	EKGK	547	
		:	:	:	:	
Db	517	HQELTFLGLLYAFTENFVLPVLSHDEVVHGKSLIAKMSGDDWQKFNALRSYYGFMWGYFGK	576			
QY	548	RLFFAG-----	-----	LRE--MYCP-RALPSDANFTMRMSYG-SIKGYEQD	585	
		:	:	:	:	
Db	577	KLIFMGQEFNAQWSENSEKSGSLDWNLRQYPMHGMGMRRLVLDNLITYRSKALHARCEPDG	636			
QY	586	GAW--YNYHTTGKVLKQDPKSDEFAVOENILDLFRTKNY-----	GRYAENGOLHIAP	637		
		:	:	:	:	
Db	637	FRWLVDHDEHNSVFAWLRTAPGKPVAVICNLTPVRYENYVPLPVAGRWRE-----	688			
QY	638	LSNND--ITGNSGSPVFDKNGRLIGLAFDGNWEAM-----	SGDIEFPD	680		
		:	:	:	:	
Db	689	ILNTDAEIIYG-SGK---GNGGRVQAVDAGGEIGAMLVLPPLATIMLEPE	734			
RESULT 8						
S44784						
C30C11.4 protein - Caenorhabditis elegans						
C:Species: Caenorhabditis elegans						
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 29-Sep-1995						
C:Accession: S44784						
R:Favello, A. D.						
submitted to the EMBL Data Library, January 1993						
A:Description: Sequence of the C. elegans cosmid C30C11.						
A:Reference number: S44782						
A:Accession: S44784						
A>Status: preliminary						
A:Molecule type: DNA						
A:Residues: 1-776 <FAV>						
A:Cross-references: EMBL:L09634; NID:g156220; PIDN:AAA27967.1; PID:g156223						
C:Genetics:						
A:Introns: 36/2; 145/3; 255/1; 417/2; 732/3						
C:Superfamily: heat shock protein 91						

```

QY 193 VFSD-----VRMYFAPSPSSVGKGGDTDNWMPRHRTGDFS--VFRVYAGADNRPAEYSKD 245
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 396 AIKDTPQYRIRLSW--NSTGNGENDVFS-PRDEVFSGKLVSLFSGPFNVFAHYAQP 451
QY 246 N-KPYKPVYFAAVSMOGYKADDYAMTIGPGSGTDRLVTSWGVEDRIENENNPRIEVRGK 304
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 452 NVPHNQVHIGSKVNGAR-----PGAD-----GGNQKV-----KVKVRYNP 488
QY 305 QGIWKEAMSAQATRIKYASKYAQSANYWKNKSIGMNRGLARLDVIGKRAEERAF--ADWI 363
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 489 DGIFTIA-----SATMVEPRIVEVPAAEMEVDCUAKTEAPAEPLPVP 531
QY 364 RKNGKSAVYGDVLSLEKAY-----KEGAKANREMYTLSETLFGGTEV 406
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 532 KKTKLVPDVLEVIESTIPVSDVQVFNHLELQMOESDAREKAKADAKNS-LEEVYVEMRK 590
QY 407 V--RFAQFANALATNPDAHAGILKSLDDKYDKYLPSLDRKVLPAMLDIVRRRIPADKLPD 464
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 591 VSDQYAEFTPAA--DEFKSVLTSTEDWLYDGBDAER-----DVEYKRL--SELKA 639
QY 465 IFKNVIDKFKGDTTKYADFVFKSVV---PYSD-----KPHAMLE$MDKKEKFAKA 514
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 640 VGTVPVVERESETRKPAPDFSDQSIMVRKAYEDYANGSGPTVAHLDSKEMEKVINAIE 698
C:Accession: F72308
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:R: Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic-
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
A:Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: F72308
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1289 <ARN>
A:Cross-references: GB:AE001761; GB:AE000512; NID:g4981529; PIDN:AAD36071.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0992

Query Match 3.3%; Score 123; DB 2; Length 1289;
Best Local Similarity 18.1%; Pred. No. 11;
Matches 125; Conservative 111; Mismatches 256; Indels 200; Gaps 29;

QY 102 DHYDLRGGFVSRTMGEEPLIPGLSVKYLKRVIVKTKQVKGQLKGITD----- 148
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 22 DSDALR--YRLNTISE-----LNKSVLKAIONLIVG-YNGLFKSGSDFFLETLYSNDPLL 72
QY 149 -----EMERLRKAQVCEQLAKENADENOLCTVEPFYNNYEFL----- 188
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 73 ARYOKESQEAIRKVOEARREGASNEETLKLQAYQEARLRENKQFIALIESNLNSIDYL 132
QY 189 -----IYDVDFKDVRFVMPAPPSSVKGCGDTDNWMPRHRTGDFS--VFRVY 233
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 133 ASETLKNLSLTANKIIDDVLNKNKEILKPLE-GDFGVTKNAILQLIDSTYKITFDYC 191
QY 234 ---GADNRPAEYSKDNKPKPVYFAAVSMOGYKADDYAMTIGPGSGTDRLVTSWGV 287
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 192 KANYGATEKTAQY-----WVKTYF-----VEPPSESEKKVL-----DEVLFKASDELK 235
QY 288 DRIENENNPRIEVRGIKOG-----IWKEA--MSAQATR-----IKYASKYAQSA 330
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 236 DRKDRUTERLKLIVSKGNALAEKVIEETAKKSAESILNIITETPSLIVELFTKYNNV 295
QY 331 NY--WKNSIGMNRGL---ARLDVIGRKRAE-----EFRAFADWIRKNGKSAVYGDVLSL 379
   || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 296 DFOITFDNDIASNEVVFYKQIRRELNGDENEINRCYFRAYFLSKKKAAKKAALNDYGERK 355

```

Qy 380 EKAYEKGAKANREMYTLSETLFGTGEVVRFAQFANAATNPDAHAGILKSLDDK----- 433  
:  
:  
D 356 QOGSTGKGRKSSEEDVIPQEVIKAEEIV-----LAKDPFLKVSAVONITRIEEMND 407  
:  
:  
Qy 434 --KYDLPLSDR-----KVLPMALDIVRRRIPADK 461  
:  
:  
D 408 LSLRALPOLNSAKRLIEYTFGOLENNEITLGEYNEYINRIVSVYTGLNTAKTKIANNL 467  
:  
:  
Qy 462 LPDIKNVI-----DKKEGDTKKYADFVF-----DKSVVPYSDDKFHAMLK 503  
:  
:  
D 468 ESEYHKGSISQSEYSEKLQLDKOVENIIQLNDFVYLKHQEIQRORLFREKLSQLVK 527  
:  
:  
Qy 504 MDKERPAKAIERDPVELSKSVIAAARAIAQADAMANAAYAIEKGRRLF---FAGLRMEYPG 560  
:  
:  
D 528 LDQSEMKL-----LDSSLA-----RFNQFLKLYKQIQAVWAFSFSFSEMMEG 574  
:  
:  
Qy 561 RALPSDANTMRMSYSGIKGYEPDQGAWNYHTTG-----KGVLEKODPKSDEF-A 610  
:  
:  
D 575 FTNGICKAFMKELNSPSPFFLLQNGWIIENLLAICQEDRLTRALIEKVDKLSSEEIKA 634  
:  
:  
Qy 611 VOENTILDLPRTKNG--RYAENGOLHIAFLSN 640  
:  
:  
D 635 LYESPICYTDEDDLVSYSKNAMLTIEEFKSS 666  
:  
:  
  
RESULT 10  
AC1647  
Lactobacillus phage phi-gle minor capsid protein 1608 homolog lin1716 [imported] - List  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1647  
R:Glaser, P.; Franqueul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluerter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A.: Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; PMID:21537279; PMID:11679669  
A:Accession: AC1647  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1571 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CAC96947.1; PID:g16414203; GSPDB:GN00178  
A:Experimental source: strain Cl1p1262  
C:Genetics:  
A:Gene: lin1716  
  
Query Match . 3.3%; Score 122.5; DB 2; Length 1571;  
Best Local Similarity 18.7%; Pred. No. 16;  
Matches 138; Conservative 104; Mismatches 308; Indels 187; Gaps 31;  
  
Qy 11 GAALLLGASGVAKAD-----KGMMLNELNQENLDNRMLGTLPDLSLYSFDKPS 61  
:  
:  
D 559 CGWLPGGGKKGKVKFLFKSKATAGLKGKFGKL----ASGVGVLCVGLSATLIGMNNKN 614  
:  
:  
Qy 62 IANAVVIFGGCGTGIIVSDOGLIFTNHHCYGA-----TQSQT 100  
:  
:  
D 615 AGEKTGGFGGSLGG--MAGAAIGTMITAPGVGTGAIGAAGTAGTALGKELYVQEGP 672  
:  
:  
Qy 101 VDHYDLRDGFVSRT-MGEELPIPLSVKYLKRIVK-----VT 136  
:  
:  
D 673 KILDRFTGWKGLSKTAEEHPILGANINVINKTIIDAAKKGIKAVGDTHKTVMNAKSIVA 732  
:  
:  
Qy 137 D--KVEGQLKGTIDMERURKAQEVCOELAKKENADENQL-----CIVEFPFYNNYF 187  
:  
:  
D 733 DPLKIDASSKGVSKD-----SAKAMNEYLRNEQMDSRVIMVSGRRIRTELENN--- 784  
:  
:  
Qy 188 LIVYDVFKDRVWFAPPSSVGEGGTDNMWPHTGDFSVERVYAGA-----DNPPAEYS 243  
:  
:  
D 785 IKTYDKMSD-QLIATAATEKKSAAKND-----WDKLVALGAVSKDTASAKKHTTNNETASIN 838  
:  
:

Qy	244	KDNKPYKPVFAAVSMOGYKADDYAMTIGPFGSDTRYLTWSGVEDRIENENNPRIEVRGI	303
Db	839	KADIKANNEELKALERBFYK-----GQEIIRNAEN---RINAI	874
Qy	304	KOGIWEK--AMSADOATRIKYASKYAQSANYWKNISGMNRLGLRD--VIGRKRAEERAF	359
Db	875	KOKASKEKRKUTKDEEKEIQ-----RMEITTTLEFRKRSERY	911
Qy	360	ADWIRKNGKSAVYGDVLSSLEKAYKEGAKANREMTYLSLTFGGTEVVRFAQFANALAT-	418
Db	912	EKEVRK-----IEBQRKEAIALTASAKEQKIILGNLENSKEKMSAKAASV	959
Qy	419	---NPDHAGHLKSLDDKYKDYLPSLDRKYLPAWLDIVRRRIPADKLPDIFKNVIDKKFK	475
Db	960	VKNSAKARDASVKEANKYKQTKKILDEK-----RFTVGEISPEEYQDALKNA--KKKK	1011
Qy	476	GDTKKYADFVEDKSVVPYSDFKHAMLKSMDKEKFAKAIERKDPAYELSKSVTAAARATQAD	535
Db	1012	NGVVKEAEKMHNDVVRQAQQAKGHLREVDWET--GETLNK--WQOFKKGTVETFNSIKDA	1068
Qy	536	AMANAAYATEKCKRFLFFAGLRWMPGCRALPSDANFTMRMSYSGIKGYEPQDCAMNYIIT--	593
Db	1069	ALGNWELSSATVAIFSQMKDT--ALRVEDSFKNLYKVINNVITGVN--KVLSPFNKITIP	1126
Qy	594	-TGRGVLEKOPKSDPEFAVQENILDLFRTKNYGRYAENGQLHIAFLSNNDITG----GNS	648
Db	1127	VLNGSIGSS-----QENLSAKDKKTYHSTQSGLNLMNMTGSGNNASGQIMAGEE	1177
Qy	649	GSPV-FDKN---GRIG 661	
Db	1178	GFEIAYNKNNSAQARILG 1194	

RESULT 11

AD1617

penicillin-binding protein homolog lin1477 [imported] - *Listeria innocua* (strain Cliph)

C:Species: *Listeria innocua*

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AD1617

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
 .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
 .; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001

A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of *Listeria species*

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1617

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-721 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC96708.1; PID:g16413950; GSPDB:GN00178

A:Experimental source: strain Cliph262

C:Genetics:

A:Gene: lin1477

Qy	277	TDRYLTWSGVEDRIENENNPRIEVRGIKQG-----IWKEAMSADOATRI-----	320
Db	123	TDRDLKDYWLTHQTESLNRLSAKEQALESSKAYKIQEKVTNDIAKLTEDLKVATYI	182
Qy	321	KYASKYAQSANYWKNISGMNRLGLRDVIGRKRAEERAFADWIRKNGKSAVYGDVLSSL	379
Db	183	KKMTTGTAMTESVINKNDVTDDEIARSENMDSLPGVDVTTTDMNRYTYDTELRSILGSV	242
Qy	380	EKAYKEGAKANREMTYLSLTFGGTEVVRF--AQFANALATNP-----DAHAGTLKS	429
Db	243	STA-KEGLPKDKAEYLYSQGYRMDRVCKSVYLEAQYESVLGAGSKSQSESULDCKGNIET	301
Qy	430	LD----DKYKDYLPSLDRKVLPAWLDIVRRRIPADKLPDIFKNVIDKKFGDTRKYYA-DF	484

Query Match 3.3%; Score 122; DB 2; Length 721;

Best Local Similarity 22.18; Pred. No. 5.5;

Matches 112; Conservative 69; Mismatches 195; Indels 130; Gaps 24;

Db 302 VSKYEGSKDLVSDVDFOKAVEDILRNKIK-----QCKQYAGSD 343  
QY 485 VFDKSVV-----PYSDKFHAML--KSMDEKFAKAIEKDPVAVELSKSVIAAARAIOAD--- 535  
Db 344 LFDRAEFVAMPYSGEVLAKGKNDKGEF-----DVSLSGTFTTAYANGSAVKGSTIL 398  
QY 536 -----AMAN-----AYAEKGRUFF-----AGLREMPGRALPSDAN---FTMRM 573  
Db 399 GGIMOGAISNKTFTDOPTALKGTPKSSWFNRTGAGNRPDPVGALEISNSSNMYQVAM 458  
QY 574 SYCSIKGYEPQ-----DCAWNYHTTGKVL-----EKODPKSDEFAVOENIL 616  
Db 459 KMGGA-KYFNGPLRAPLSTFDDMRYYYNOFGLGVTGIDLPGEQGYGDDGTICK-IL 516  
QY 617 DLFRTRNGRYAENGOLHIAFLNSNDITGNSGSPVFDK-----NGLRIGLAFDGNWE 669  
Db 517 DF-----AIGQYDSTPLQMAQYVSTTANGSRIAPSMVKREINPSTNGDSVGLTAN-- 570  
QY 670 AMSGDTEFPDLQRTISV---DIRYV 692  
Db 571 -----BPKVLNKGVSNDIKTV 588  
  
RESULT 12  
DNA helicase II U0501 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: D82883  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min  
A:Reference number: A82870  
A:Accession: D82883  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-743 <GLA>  
A:Cross-references: GB:AE002148; GB:AF222894; NID:g6899495; PIDN:AAF30913.1; GSPDB:GN001  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: uvrD; U0501  
A:Genetic code: SGC3  
C:Superfamily: helicase II  
  
Query Match 3.3%; Score 122; DB 2; Length 743;  
Best Local Similarity 17.7%; Pred. No. 5.8;  
Matches 143; Conservative 129; Mismatches 281; Indels 256; Gaps 40;  
  
QY 10 LGAALLGASGVAKAD-----KGMWLLNELNQENLDRMRELGFLL---PLDSLYSFDKPSI 62  
Db 22 LGPOLVIAGAGTGKTSVLTIRIAYLITE---KNHPSRILGFTFTNKADEM---KERV 74  
QY 63 ANAVVI-----FGGCGTCTVSDGLFTNHHCGYGAQSOSTVDHDLRDCGFSRTM 115  
Db 75 GKTIGVSIPLYSTFHSWC---VKLIQDDIHLNLYH---NNIKIIDDDQEVL-----LKEI 124  
QY 116 GEELPIPLGSVKYLRIVKTVKVEQLAGITD-----EMERLKAQEV----- 159  
Db 125 FDLNIEKKS-QVKKIITISKVKNKFFDNDMLNEKNHYLELVDLNDQAORLVDIYKI 193  
QY 160 -COELAKKENADENQIC-----IVEPFYNNSEYFLIYVDFKDV- 198  
Db 184 YCDRCFKLVYVDFDDLINLTHKLFIEFFVELEKWNKFDYILV--DEFQDTNKIYDLIS 241  
QY 199 -----MYFAPSSVKGKGGDDTNMMWPRHTGDFSFRVRYAGADNRPAYSKDN-K 247  
Db 242 LIATKHQNLVGVDPQOMIYSFRG-AEQWLIINNFQNEK-----NTKVTILKTNYR 291  
QY 248 PYKPVYFAAVSMOGYKADYAMTIGFPGSTDYLTLSWGVEDRIENENNPRIEVRG---IK 304  
Db 292 STOPILNTANRLIDANNY-----KKNLTAFNT-----NDNNLPILRGQNPID 336

QY 305 QGTW-----KEAMSDAQATRIKAYSKYAQSANYSKNSIGMNRGLARLDVIGKRRAE 355  
Db 337 EANNIARKIRELLEEGTPANQIAVLFNSNHYSTIEQ-----SMRRESIPYTLILGSKKFY 391  
QY 356 ERA-FADWIRKNSAVYGDVLSLSLEKAYKEGAKANREMYL-----SETLFGTEVVRF 409  
Db 392 ERAEIKDMI-----AYLKVVNDLD-----ELSFRLIINTPRAIGTPTFEHV 433  
QY 410 AQPA-----NALATNPDAH-----AGILKSLDDKYKDYLPSLDRKVLPA 448  
Db 434 KHVAIINNLFLEALAEVEKNHLINNTOKKNILNFVNLIKEIRDEMEL-----KIHE 486  
QY 449 MLDIVRRRI-----PADKLPIFNVDKFKGDTTKYADVFVKSVVYSDKPHAML 501  
Db 487 ILELIYKKNVYAYLLENEKAEDKIDNVE--LKRAMKMYVDRHPDDTINDYLSIALYL 544  
QY 502 KSMDEK-----FAKAIKDP--AVELSKSVIAAARAIOADAMANAYAIKGRKL 549  
Db 545 NKDGKOSKENVLLMTVHNSKLEYENFVAGMNEGLLPSDRAINDDPIK---GVEEERRI 601  
QY 550 FFAGL-----REMPGRALPSDANFTMRMSYSGSIKGYEPQDGAWY 589  
Db 602 AYVALTRAKKNLYISSACCYDPLARRQVFESFINEIGFNNLKIINSSFKNNKPED---- 657  
QY 590 NYHTTGKGVLEKODPKS--DEFAVOENILDLFRTKNYGRYAENGQLHIAFLNSNDITGNN 647  
Db 658 ---MPLKSLFKQOERSWFSKQKQKQEVEDNFYQ-----TMKND----- 693  
QY 648 SGSPVFDKNGRLIGLAF-DGNWEAMSGDI 675  
Db 694 -----FEIGERIVHTSFGDGVIIIGDGI 717  
  
RESULT 13  
C95066  
valyl-tRNA synthetase [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
C:Accession: C95066  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H  
nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: C95066  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-883 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK74724.1; PID:g14972043; GSPDB:GN00164; TIGR  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0568  
C:Superfamily: valine-tRNA ligase  
  
Query Match 3.3%; Score 122; DB 2; Length 883;  
Best Local Similarity 20.0%; Pred. No. 7.5;  
Matches 102; Conservative 66; Mismatches 167; Indels 176; Gaps 24;  
  
QY 192 DVF-----KDVRYVFAVPPSVGKGGDTNNMWPRTHTGDFSFRVRYAGADNRPAYS 243  
Db 26 DVEKPSGDQAKPYSIVIPPNVTGKL-----HLG-----HA 57  
QY 244 KDKPKYKPVYFAAVSMOGYKADYAMTIGFPGSTDYLTLSWGVEDRIENENNPRIEVRG 302  
Db 58 WDT-TLQDIIIRKRMQGF-----TLWLPQMDHAGIATQAKVEERLREGITRYDLGR 110  
QY 303 ----IKOGIKAMSDAQATRIKYA-SKYAQSANYSKNSIGMNRGLARLDVIGKRRAE 357  
Db 111 ESFLTKVWEKND---EYATTIKEQWCKMGLSVDSYRERETLDEGLS-----KAVRK 158  
QY 358 AFADWIRK-----NCKSAVYGD-----VLSLEKAYK-----EGAKANREMT 394

159 VFVNLKGYRGEFTINDDPAARTALSDEIVHKDVEGAFYHMNYMLEDGSRALFVAT 218  
395 YLSETPGGTEVVVFAFANALATNPDAHAGILKSLDDKDYKVLPS-----LDKVLPLAM 449  
219 TRPTEMGDV-----AVAVNPE-----DPRYKDLGKNVILPANKLIPIV 259  
450 LD-----IVRRRIPADKLPDIEKNVIDKKFKGDTKKYADFVFKSVVPYSDK 496  
260 GDEHADPELGTGVVKITPAHPNDPLVGRHNLPOVNMDDGTMLNELAFESCM---DR 316  
497 FHAMLKSMDEKFAKATEKDPVLSKSVIAAARA-----IOAD 535  
317 FEAR-----KAVVAKLEIGALVKIEKRHSVGHSGERTGVVPEPRLESTQFWFKMDQLAKN 371  
536 AMANAYAIKGRKLPFAGLREMYGPRALPS-----DANFTMRMSYSGIKGYEPDQG 586  
372 AIANQDTEK-----VEFYPPRFNDTFLLOWMENVHDWISRLQMWGH-----OIP 416  
587 AWNYHTTGRGVLEKQDPKDEFVAVQENILD 617  
417 AWYN--ADGEMYVGEAPEGDGWTODEDVLD 445

RESULT 14  
D69103  
DNA helicase (EC 3.6.1.1) MTH1770 [validated] - Methanobacterium thermoautotrophicum (str. N) alternate names: cell division control protein CDC21/CDC54 homolog; replication license N; Contains: adenosinetriphosphatase (EC 3.6.1.3)  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 26-May-2000  
C:Accession: D69103  
R:Smith, D.R.; Doucet, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Olu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N. J. Bacteriol. 179, 7135-7135, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: function N; Reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: D69103  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-666 <MTH>  
A:Cross-references: GB:AE000932; GB:AE000666; NID:g2622894; PIDN:AAB86236.1; PID:g262290  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH1770  
C:Complex: dodecamer; double hexamer [validated, MUID:20144074]  
C:Function: <HEL>  
A:Description: EC 3.6.1.1; DNA helicase; ATP-dependent DNA unwinding activity [validated]  
C:Function: <ATP>  
A:Description: EC 3.6.1.3 [validated, MUID:20144074]; adenosinetriphosphatase, DNA-dependent  
C:Function: <DNA>  
A:Description: ATP-independent single-stranded DNA-binding; ATP-dependent double-stranded  
C:Superfamily: Archaeoglobus replication licensing factor MCM-type homolog; MCM homology  
C:Keywords: cell cycle control; DNA replication initiation; hydrolase  
F:98-570/Domain: MCM homology <MCM>

Query Match 3.2%; Score 120; DB 1; Length 666;  
Best Local Similarity 19.3%; Pred. No. 6.6;  
Matches 116; Conservative 94; Mismatches 232; Indels 158; Gaps 28;

119 LPIPLGLSVKYLK-----IVKVTDKVEGQL-----KGITDEMERLRKAEV 159  
96 IPLRLRSKFTGKFAVVDGIVRKTDIEPRIVKAVFECRCMRHHAFTQSTNMITE-PSL 154  
160 COELA-----KKENADENQICIEPPE--YSNNFYFLIVDFKDVVRVAPPSSVG 208  
155 CSECGGRSFRLLQDESEFLDTQTLKLOEPLNLSGGGQPRQITVVLDEDDLVDTLTPGDIV 214  
209 KFGGDT-----DNNWPRHT-----GDFSVFVYAGADNRPAEYKDNKPY-KPV 252  
215 RVVTGLTVRDERTKRKNFYGNVTEFLQGFPELQISEDEKIKELAGDPNIYEKII 274

253 YFAAVSMQGYKADDYAMTIGTGGSTDRYLFSGWGVEDRIENENPRIEVRGKIGKIMKEAM 312  
275 RSTAPSIHGYREVKEALALQLFGGTGKEL-----DDKTRLRGDIHLIVG-DPGIGKSQM 328  
313 SADQATRIKASYKAQSANYWK-----NSIGMNRGLARLDVIGRKRAERAFADWIRKNGK 368  
329 -----LKTVSKLAPRGITVSGTSGVGLTAAVVRDE-----FGWSLEAG- 369  
369 SAVYGD-----VLSSEKAYKEGAKANREMYLSLSETLFGGTEVVVRFAOFANALATNPDAHA 424  
370 ALVLGDKGNVCVDELDMKREDSAIHE-----ALEQQTISTAKA 409  
425 GILKSLD-----DKYKDYLPSLDRKVLPA-----MLDIVRRRIRPADK 461  
410 GIMATLNSRCSVLAANPKFGRFDSYKIAEQID---LPSTILSRFDLIEFVEDKPDDEK 466  
462 LPDIFKNVIDKFKGDTKKYADFVFKSVVPYSDKFIAMLKSMDEKFAKAEKDPVLE 521  
467 DRELARHIL-KTHKED---HMPFEIDPELL---RKYIAYAR---KNVRPVLUTDEAMQV 514  
522 SKSVIAAARAIOAD---AMANAYAIKGRKLPFAGLREMYGPRALPSDANFTMRMSYGS 577  
515 LEDFYVSMRASADESDPVPITARQLEALVRLSEASAKIKLKEHVEAEDARKALKLSQAC 574  
578 IK--GYEPQDGAWNYHTTGTGVLEK---QDPKS---DEFVAVQENILDFTKKNYGRYAE 630  
575 LKQVGYDPE-----TGKIDIDKVEGRTPKSERDKFRLLLELLEIKEYEDDYGGRAPTN 625

RESULT 15  
C82332  
translation elongation factor EF-G VC0361 [imported] - Vibrio cholerae (strain N16961)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: C82332  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, L.R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000  
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: C82332  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-698 <HEI>  
A:Cross-references: GB:AE004124; GB:AE003852; NID:g9654770; PIDN:AAF93534.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC0361  
A:Map position: 1  
C:Superfamily: translation elongation factor G; translation elongation factor Tu homo  
F:11-145/Domain: translation elongation factor Tu homology <ETU>

Query Match 3.2%; Score 118; DB 2; Length 698;  
Best Local Similarity 18.4%; Pred. No. 9.5;  
Matches 120; Conservative 83; Mismatches 202; Indels 246; Gaps 26;

8 ILLGAALLL-GAGVAKADKGMW-----LLNELNQENLDRMRELFTLPDLSD 54  
105 VLDGAVVVGCTGCTVETVVRQADKYGVPRMVFYKMKDRACADFLRVVG----- 156  
55 YSFDPKPSIANAVVI-----FGGCTGTV-----SDOGLIFT----- 86  
157 -QIKHRLGANPVPIQLNIGAEFEFGVIDLKKMAKNNEADQMSFTYEIPADMLELA 215  
87 ---NHHCGYGAIQSQSVVDHYLRDPVS-----RTMGEELPIGLSVKYLKIV 133  
216 QEMRNLHIVAAAEASEELMEKYLEDGELSEVEIKOALRQRTINNEIVLAACGSFAKKNKV 275  
134 K-VTDKY-----EGQLKGITDEMERLRKAEQVCEQELAKKENADENQICIEPPEYSN 183  
276 QAVLDVAVIEFLPSPTDVPATKIGIDD-----RENSVERHADDN-----EPF--S 316



```
QY 184 NEYFLIYDVFKDVRMVPAPSSVCKFGGDTDNWMPRHGTGDFSFRVYAGADNRPAEYS 243
Db : | | | |
317 SLAFKIATDPF-----VGSLEIRIYSGVWNSGDVY 348
QY 244 KDNKPKPYFAAASMOGYKAD-----DYAMTIG----- 272
Db : | | | |
349 NSVKQKKEFRGRIYQMHANKRDEIKEIRAGDIAAAGLKDVTTGDTLCDPNHVVILERME 408
QY 273 FPGSTDRYLTSGVEDRIENENNPRIEVRGKQGIWKEAMSA--DOATRIKYASKYAQS-- 329
Db : | | | |
409 EP-----EPVQIAVEPRSKADQEKMGIALGKLAEDPSFRVETDAETGQILI 456
QY 330 -----ANYWKNISIGMNRGLARLDVIGRK--RABERAFADWIRKNGKSAVYG--- 373
Db : | | | |
457 SGMGELHLDIIVDRMKREFGVDCNVGKPQVAYRETIRGKSEVEGKFRVQSGRGQYGHWW 516
QY 374 -----DVLSSLEKAYKEGAKANREMTYLSSETLFGGT 404
Db : | | | |
517 LKIEPAEPGQGFVFDATAGGVIPKEFINPVAKGIEEQMNGVLAGYPVLDVKATLFDGS 576
QY 405 -EVVFAQAFANALATNPDAHAGILKSL-----DDKYKDYLPSLDRK--VL 446
Db : | | | |
577 FHDVDSSEMAFKIAGSMAFKKGALEAQPVLLPLMKVEITTPEDWMDVGVGDLNRRRGII 636
QY 447 -----PAMLDIVRRRIPADKLPIDFKNVIDKK-----FKGDTKKYAD 483
Db : | | | |
637 EGMDEGPAGLKIIHAKVP---LSEMFYATDLRSATQGRASYSMEFAEYAD 684
```

Search completed: May 25, 2003, 15:34:45  
Job time : 53.9435 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2003, 15:31:51 : Search time 52.8284 Seconds  
(without alignments)  
2777.023 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MQMKLSILGAALLGASG.....LFMDKWGQCPRLIQELKLI 712

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	907.5	24.4	716	16 Q9PC94	Q9pc94 xylella fas
2	146.5	3.9	1368	16 Q992W2	Q992w2 streptococ
3	133.5	3.6	1104	2 Q06812	Q06812 micrococ
4	131.5	3.5	991	17 Q82VU0	Q82vu0 pyrobaculum
5	130	3.5	1175	16 Q8RHT3	Q8rht3 fusobacteri
6	126	3.4	953	2 Q48502	Q48502 lactobacill
7	125	3.4	1440	5 Q9TYU6	Q9tyu6 caenorhabdi
8	124.5	3.3	735	16 Q8U8L4	Q8u8l4 agrobacteri
9	124	3.3	1808	10 Q9LQJ8	Q9lqj8 arabidopsis
10	123.5	3.3	713	2 Q85051	Q85051 moraxella c
11	123	3.3	1289	16 Q9X087	Q9x087 thermotoga
12	122.5	3.3	742	2 Q9ZB47	Q9zb47 streptococ
13	122.5	3.3	1571	16 Q92B35	Q92b35 listeria in
14	122.5	3.3	2204	12 Q99FL1	Q99fl1 porcine tes
15	122	3.3	721	16 Q92BR7	Q92br7 listeria in
16	122	3.3	743	16 Q9PPY9	Q9ppy9 ureaplasma

#### ALIGNMENTS

#### RESULT 1

Q9PC94 PRELIMINARY; PRT; 716 AA.  
ID Q9PC94  
AC Q9PC94;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein Xf1887.  
GN Xf1887.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,  
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfi H.,  
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.B., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
RA Menck C.F.M., Miraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

Q97s45 streptococ  
Q8tqb8 methanosarc  
Q9llp5 oryza sativ  
Q8t5c7 plasmodium  
Q27798 methanobact  
Q8rt51 fusobacteri  
Q99ysl streptococ  
Q99fl7 porcine tes  
Q06915 lactobacill  
Q9gsk2 chilo iride  
Q92215 c putative  
Q92k97 helicobacte  
Q8rd19 thermoanaer  
Q8wz53 homo sapien  
Q9kb24 bacillus ha  
Q9y059 caenorhabdi  
Q9y059 caenorhabdi  
Q47766 enterococcu  
Q9kc46 bacillus ha  
Q22830 caenorhabdi  
Q51284 neisseria m  
Q9xcv8 lactobacill  
Q9jtf1 neisseria m  
Q8wrf8 toxoplasma  
Q8srk4 encephalito  
Q00936 toxoplasma  
Q8wrf9 toxoplasma  
Q8ryng9 oryza sativ  
Q977z3 thermoplasm  
Q92291 listeria in

da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A., de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H., Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., Zago M.A., Zatz M., Weidman J., Setubal J.C.

RA The genome sequence of the plant pathogen *Xylella fastidiosa*.  
RT Nature 406:151-159(2000).  
RL EMBL: AE004008; AAF84693.1; -.  
DR MEROPS; S46.001; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BBDFC CRC64;

Query Match 24.4%; Score 907.5; DB 16; Length 716;  
Best Local Similarity 31.4%; Pred. No. 9.9e-51;  
Matches 239; Conservative 135; Mismatches 291; Indels 95; Gaps 21;

QY 1 MOKKLKILGALLGASGAVAKADKGMWLLNELNQENLDRMRELFTLLPLDSLISYSPDKP 60  
DB 1 MRFNLLSLVALITVDS--THAGEGMWVQQ--PIAGPLQAQGLQSPQLSNITGD 57  
QY 61 SIANAVIFGGGCTGTVSDQGLFTNHHCYGAIQSQSTVDHLYLDRGFSVRTMGELP 120  
DB 58 PMGAVVSL--GNCTASLVSPGLVITNHHCAYGAIQLNSTPKKNLIKEGFNALTQADEV 115  
QY 121 I-PGLSVKYLKIVKVTQKVEGQLKGTIDEMERLKAQFVC--QELAKKENADENQLCIV 177  
DB 116 AGPNARIYVLEQITDVTQAQKALAAAGNDPFRKTTTLETFSKQEIACBEE-EGGYRCQF 174  
QY 178 EPFYSNNEYFLIVYDVKQVRYFAPPSSVYKFGGDDNMWPRHTGDSFVRYACADN 237  
DB 175 FSPAGGWTYRVFKMLEIKDVLRYVAPOGSVKFGGDDNMWPRHTGDSFVRYAVKGDG 234  
QY 238 RPAEYKDNKPKYKPYFAAYSMQGYKADYAMTIFGPGSTDRYLTSGWEDRIENENNR 297  
DB 235 KPASFSENIPYRPHKLFESDQPLGDSFVWVAGYPCRTNRYALV-----AEFENTAH 288  
QY 298 IEVRGIQ-----GIKWEAMSDQATRIKAYSKYASANYWKNKSGNMRGLARLDVIGR 351  
DB 289 WTYPIVIGQHFKNLJALTEAASKQNPDIQVYASTLAGLNNTSKNFGDQLDGRFRINAIG 348  
QY 352 KRAPERAFADWIRKNG---KSAVYGVLSLEKAYKREGAKANREMYLSETLFGGTEV 407  
DB 349 KQSEETAVLAWLKQOIRGHEALAAHQTLVDLTQY---KANQDRDFVLQ--FNGSGVI 403  
QY 408 RFAQAFANALA--TNPDH-----AGILSKLDDKYDYLPDLDRKY-----445  
DB 404 GVAVNLYRLALERTKSDAQRAGYQERDLPTIEGNLKQME---RRLPEMDROMQOYWLT 460  
QY 446 ---LPALMDIVRRIPA-----DKLPDIFKNVIDKKFKGDTTKYADFDVSKSVVPSD 495  
DB 461 EYNKLP-----VKORVAADVWLGDIPTATLRLGDTKLSSSEER-----500  
QY 496 KFHAMLSMDKEKFAKATEKDPAVELSKSVIAAARAQAQAMANAYATEKGRKLFAGLR 555  
DB 501 -----LKFNFADRAAFESSQDPATRYAVAINPALLETIONKIRTEGELLKARPIYLOALA 555  
QY 556 EMYP--GRALPSDANFTMRMSYSGIKYEPDQGWYHTTGKGVLEKQDPKPSDEFVQVE 613  
DB 556 DYNKSHGKFVYPDANSSLRITFGHVKGYSKPDGVEYPTFTTQQGVMAK--NTGVPEFDSPK 614  
QY 614 NILDLFRTKNYGRAEN--GOLHIAFLSNNDITGNSGSPYDFKNGRLIGLAFDGNHEAM 671  
DB 615 SLINAIKAKSYANLADQRIQVVPVNFSLDLDITGNSGSPVLDAGKLVGLAFDGNMESV 674  
QY 672 SGTDFEFLDQRTISVDIRYVLFMDKWCQCPRLIQELK 711  
DB 675 SSNWVDFPVTTRTAVDSRYVQWIMTEVAPAPHLKELNL 714

RESULT 2  
Q992W2 PRELIMINARY; PRT; 1368 AA.  
ID Q992W2  
AC Q992W2  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Hypothetical protein SPY1046.  
GN SPY1046.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;  
RX MEDLINE=21192684; PubMed=11296296;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;  
RA "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*."  
RT Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
RL EMBL: AE006549; AAK33936.1; -.  
DR Hypothetical protein; Complete proteome.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1368 AA; 158440 MW; 07D04F0B5965762F CRC64;

Query Match 3.9%; Score 146.5; DB 16; Length 1368;  
Best Local Similarity 19.5%; Pred. No. 1.3;  
Matches 167; Conservative 105; Mismatches 228; Indels 357; Gaps 43;

QY 35 NQENLDRMRELFTPLDSLISYSPDKSIANAVIFGGCTGTVSDQGL--IFTNHHCY 92  
DB 609 NEENEDILEDIVLTLTFE---DREMIERLKYAH-----LFDDKVMKQLKRRRTGW 659  
QY 93 GAT-----QSQSTVDHLYL--DGFVSR-----TMGEELPIPLGLS----125  
DB 660 GLSRKLINGIRKQSGKTI--LDFLKSDGFANRNFOLIHDSLTFFKEDIQKAVSQGD 718  
QY 126 -----VKYLRKIVKVD---KVEQ-----LKGIDEM 150  
DB 719 SLHEHIANLAGSPAIIKQILQTVKVDVDELKVMGRHKPENIVIEWARENQTKQGNRSR 778  
QY 151 ERUKAQEQVQELAK---KENADEN--QLCIVPEPYSNNEYFLIVY-----DVFQDKRMVF 201  
DB 779 ERMKRIEELKELGSQLKKEHPVENTQL-----QNEKLYLYLQNGRDMYVDQELDI 830  
QY 202 APPSSVKFGGDTDNWMPRHGTGDFSVRYAGADNPAEYKDNKPKYKPYFAAVSNQG 261  
DB 831 -----NRLSDYDVH-----IVPQS 845  
QY 262 YKADYAMTIGFPGSTDRYLTSGWEDRIENENNPRIEVRGIKQIKWEAMSDQATRIK 321  
DB 846 FLKDD-----SIDNKVLTDRSKNKGSDNVPSEEVKKMKNYRQLLNKALITQRK 896  
QY 322 YASKYASANYWKNKSGNMRGLARLDVIG---RKRAERAPADWI-----RKNGK----368  
DB 897 F-----DNLTKAERG---GLSELDKAGFTKQRLVETQITKHAQILDSRMNTKYDEN 946  
QY 369 -----SAVYGDVL-----SSLEK 381  
DB 947 DKLIREVKVITLKSLSVDFPKDFQYKVRINNYHHAHDAYLNAVGVGTALIKYPKLES 1006  
QY 382 AYKEG-----AKANREMYLSETLFGGTEVVRFAQAFANALATNPDAHAGILKS-429  
DB 1007 EFVYGDYKVDYVRKMKIAKSEGEIGKATAKYFYFNIMFFKTEITLANGEIRKRPLIETN 1066  
QY 430 -----LDDKYKDYLPSLDRKVLKPL--AMLDIVRRR-----IP---ADKLPIDFK 467  
DB 1067 GETGEIVMDKGRDF--ATVRKVLVSMQVNIIVKTEVQTGTFGSKESILPKRNSDKL-----1119  
QY 468 NVIDKKFKGDTTKYADFDVSKSVVPYSDFKFA-----MLKS-----MDKEKFA 510  
DB 1120 --IARKKDWDPKYGCG--FDSPTVAYSVLVAVKVEKSKKLKSVKELLGTTIMERSSE 1175  
QY 511 -----KATEKDPAVELSKSVIAAARAQAQAMANAYATEKGRKLFAGLRREMY 558  
DB 1176 KNPIDFLEAKGYKEVKKDLIIKLPK-----YSIFELENGKRKRLASAGELQ 1221

Db	641	KLNTATLQGLGSHDEEGLISLADGKGLQVATLQATKQCPVYYG--BELGQTGA	699
QY	436	DYLPSSDRKVLPAMLDIVRRRIPADKLPIEFKNVIDKKEGDTKKYADFPVFDKSVVPYSD	495
QY	496	KFIAMLKSMDKERFAKLEKDPAVELSKSV---IAAARAIQADAMANA-YATEKGKRLEFF	551
QY	552	AGLREMYPGRALPSDANFTMRMSYGSIGK-----YEPOQCAWYN	590
QY	591	YHTTGGVLEKQPKPDEFAVQENIIDLFTKNYGRYAE-----NGQLHIAFLSNNDITGG	646
QY	647	NSGSPVFDKNGRLIGLAFDGNWAMSGD--IEEPF-DLORTTSVDIRYVLFMID-----	697
RESULT 4			
Q8ZVU0 PRELIMINARY; PRT; 991 AA.			
Id	Q8ZVU0		
AC	Q8ZVU0;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	PAR2P2b.		
OS	PAE2129.		
GN	Pyrobaculum aerophilum.		
OC	Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;		
OC	Thermoproteaceae; Pyrobaculum.		
OX	NCBI_TaxID=13773;		
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=IM2 / ATCC 51768 / DSM 7523;		
RX	PubMed=11792869;		
RA	Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,		
RA	Miller J.H.;		
RT	"Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum		
RT	aerophilum.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).		
KW	EMBL; AE009858; AAL63966.1; -.		
DR	Complete proteome.		
SW	SEQUENCE 991 AA; 111981 MW; 1C7D26FD8545138D CRC64;		
Query Match 3.5%; Score 131.5; DB 17; Length 991;			
Best Local Similarity 19.0%; Pred. No. 7.3;			
Matches 138; Conservative 110; Mismatches 244; Indels 235; Gaps			
QY	111	VSRTMGELPIP-----GLSVKYLRIKVVTKVGEOLKGIIDEM-ERLRKAQEQCOEL	163
QY	164	AK--KENADENQLCIVPEFFYSNNEY-----FLIVYD	192
QY	193	--VEKDVRMVAPSPSSVGKFGGD-----TDNNMPPRHGTGDFSFRVRYAGADN	237
QY	238	RPABYSKDNRPKPYVFAAYSMOGYKADDYAMTIGPGSTD-----	278

```
QY 279 ---RYLTSGWVEDRIENENNRIEVRGKIQG-IWKE-AMSAQ-----ATRIKY--- 322
Db 500 VKRFQIDGVDVREMFKLEEVVWMAEDVYRVNVTVERDOLDTSGQLPKTVIRFKLG 559
QY 323 ASKTAOSANYKNSGNRGLARLDVIGRKAERAF-----ADWIRKNGKSAYVGD 374
Db 560 GEEVAYITVYWTGS-----KLLAOFDG-SREKAEGLASIIITAGLGAKEYKPKNAKWTVH-- 612
QY 375 VLSLEKAYKEGAKANREMTYLS-----ETLF--GGTEVVREAFANALATNPDA--HA 424
Db 613 -----LHTDGIITAIRHDSWENAVGFVDLYSKGLISEDRIYEQLVNDIATGPTNYYIA 665
QY 425 GILSKDDKYYKDYLPSSDKVLPAMLDIVRRRIPADLPDIFKNVIDKKFKG----- 476
Db 666 GVEFSVYHKTEGKYDFLEIRYHPG-----NEASKNAVNALKARGLKEGVH 711
QY 477 -DTKYADFVFDKSVVPYSDFKHAMLSMDKEKFAKAEKDPAVELSKSVTAAARATOAD 535
Db 712 FTVKEYGDYEIHAVESYTKALEALARSWLKGEYYAID-----DRKRVISVKAEBKD 764
QY 536 AMANAYAEKGRKFFFAFLREMYPGRALPSDANFTMRMSYSGIKGYEPDQGAWYNYHTG 595
Db 765 AVVWT-----LKTAGLEP-----DRHETKWM-----DG-YTVIRITY 795
QY 596 KGVLEKQ-----DPKSEFAVOENILDFRTKNYGRYAENGOLHIAFLSNNDITGNSG 649
Db 796 EGLREIORMALNGDVEAERFI--RELEDVLR-RRHGONAVKRLIEVLTTPAREE---GTID 849
QY 650 SP--VFDKNGRLIGLADGNWEAMSGD-----LEFPDLORTISVDIRY 691
Db 850 LPLEVRERGNVAVRVVDLRYEFYMGOPVQGCAGEDCLRLIYEYETEGERR---QLKI 906
QY 692 VLFMIDK 698
Db 907 VCYWVK 913

RESULT 5
Q8RHT3 Q8RHT3 PRELIMINARY; PRT; 1175 AA.
AC Q8RHT3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN Hypothetical protein FN1912.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Wallunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2003-2018(2002).
DR EMBL; AE010493; AAL94011.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1175 AA; 133514 MW; 9848DE32166E3579 CRC64;

Query Match. 3.5%; Score 130; DB 16; Length 1175;
Best Local Similarity 19.4%; Pred. No. 12;
Matches 148; Conservative 112; Mismatches 282; Indels 220; Gaps 38;

QY 25 DKGMWLLNELNQ-----ENLDRRELQ----FTLP-LDSLYSFDKPSIANAVIFGGCT 74
Db 372 DKVEFVLENKANVEGDIKNLOGAVDGLSITITLPSPKDFVITCKSIKNSIV----NIS 427
```

```
QY 75 GITVSDOGLIFTNHHGCGATQSOSTVDHLYRDGFSVRTMG-----BELPIPLGSLVYLR 130
Db 428 GINL--DNNLITGKYNLKDKNLDLRVLSSEKHEKYAKOLGYLYGQIDVKGVAGK-IK 485
QY 131 KIVK-----VTDKVEGOLKGITDEMERLRK 155
Db 486 AVAKGRATNFEKLPDLAYDIEYNAENYSDGIASINGLIIDIDRQYDGLLVGKVRNLKEK 545
QY 156 AOEVCQ-----ELAKEN--ADENQLCIVEFPYNNYFLLIVYDFKDVVRVFPAPPSVG 208
Db 546 TLDIKKHNKIDLAKLONLJNSNPNIKIGIVNTDETIN----- 581
QY 209 KFGGDTDNMMPRITGDFS-----VFRVYAGADNRPAEYSKO--NKPKVPYVFAAVSM 259
Db 582 ---GTINN---PTKLDISSEVSISIKTFKINDISLNLTGOKEASLKNLNDVYKNLIVG 635
QY 260 QGY---KADDYAMTIGFPGSTD-----RYLTSGWVED-----RIE--NENNPR--IEVR 301
Db 636 NGYVDIKNTYNLLVKSNNKIDVSKFQSFLLTPYGIENAKGNIALNIETNENTEKYINLE 695
QY 302 GIKOGIWKEAMSADOATRIKAYSAOSANYKNSIGNNRGLARLDVIGRKAERAFAD 361
Db 696 NISL-----ESTKAK-----LKLSPF---SGPINFERRIDV--GELRAS----- 730
QY 362 WIRKNGKSAYVGDV-LSSLEKAYKEGA-----KANREMTYLSLTFGGTEVVRFAOFAN 414
Db 731 --LNNSPLVVDGFFVDLANISKDKEDLIRSLPYLKHFMDFHFN---YFPELIISGSTE 785
QY 415 ALATNPDAHAG-ILKSL-----DDKYDYLPISDRKVLPAMLDIVRRRIPADLPDIFK 467
Db 786 LTATNEEYVGNLIIDALIVYDIPNNYRDFSLIREQLRKRRTDVWSTKID----- 836
QY 468 NVIDKFKGDTKYADFVFDKSVVPYSDFKHAMLSMDKEKFAKAEKDPAVELSKS VIA 527
Db 837 ---DKQSKTDKEV-----EEMRMLNKLMPIDFVVKTEKPIILIDMDNENIV 880
QY 528 AARA---IOADAMANAYAEIKG-----RLFFAGLREMYPGRALPSDANFTMRM 573
Db 881 VPEYVGKLYIDLNLNG---KKGYYITGETELKEGYFVGTNEFQVDRAL---AVFNENV 934
QY 574 SYGSI-----KGYEPDQGAWYNYHTTGKGVLEKQDPKSEDFAVQENILDLFRTKNYGR- 626
Db 935 PLPEINPIFFESTIEMDDEEYHENTAGKYNQLRYEISSKTAKVGGDLISALIVPNNADEH 994
QY 627 ---VAENGOLHIAFLSNNDITGNGSPVFDKNGRLIGLAFD 665
Db 995 IYSYGDGNEIFITPMKN--LIAGQAGQVVFSGTTRYTKRKFPD 1034

RESULT 6
Q48502 Q48502 PRELIMINARY; PRT; 953 AA.
AC Q48502;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Alpha-amylase.
GN AMYA.
OS Lactobacillus amylovorus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Lactobacillaceae; Lactobacillus.
OX NCBI_TaxID=1604;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIP 102989;
RA Giraud E., Cuny G.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-155 FROM N.A.
RC STRAIN=NRRLB4540; PubMed=7986030;
RX MEDLINE=95077354;
RT Fitzsimons A., Hols P., Jore J., Leer R.J., O'Connell M., Delcour J.;
RT "Development of an amylolytic Lactobacillus plantarum silage strain
```



OS *Agrobacterium tumefaciens* (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Woo L.,  
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens*  
RT C58.";  
RL Science 294:2317-2323(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houmlel K., Gordon J., Vaudin M., Tarchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
RT *Agrobacterium tumefaciens* C58.";  
RL Science 294:2323-2328(2001).  
DR EMBL; AE009339; AAL44878.1;  
DR EMBL; AE008276; AAK89352.1;  
KW Complete proteome.  
SQ SEQUENCE 735 AA; 83760 MW; 988721391DCL16691 CRC64;

Query Match 3.3%; Score 124.5; DB 16; Length 735;  
Best Local Similarity 20.4%; Pred. No. 13;  
Matches 157; Conservative 83; Mismatches 255; Indels 275; Gaps 44;

Qy 108 DGFVSRTM--GEELPIGLSVKYLKRVKVDKVEGOLKGITDEMELRLKRAQVCOELA 164  
Db 43 EGFSARCIPGAEEVSVLTLDGNFVGEUKQIDP--DGFEGRIDLSKQPVRYRACR--- 97  
Qy 165 KKENADENQLCTVEPFYSN-----NEYF-----LIVYDFKDVDM----- 199  
Db 98 -----DDAEWAVTDYSGPVLGPMDDYFVREGSHLRLED-----RMGAHPKLEGVGEF 147  
Qy 200 ---VFAPSSVKGFGDNDNMWPRH-----TGDFSVFR--VYAGADNRPAYSKDNK 247  
Db 148 HFAVWAPNARVSVVGFDFNNDDRRHVRFRKDTGIWEIFAPVYAGC----- 195  
Qy 248 PYKPYFAAVSMQG-----YKADYYA-----NTIGFPSTDRYLTLSWGVEDRIENEN 294  
Db 196 AYK---FEILGANGELLPLKADPVPARRGELRPKNASVTAPELTQKW-----EDQAREH 246  
Qy 295 NPRIEVR-----GIKQIGKEA-----MSADQ--ATRIKYASK----- 325  
Db 247 WAQVDORQPIISIEVHAGSWQRREDGTFLSWDELAALQILPYCTDMGTHIEFLPITEHP 306  
Qy 326 -----YAOSANYKNSIGMNRGLARLDVIGRKRAEERAFADWI----- 363  
Db 307 YDPSMGYOTTLGYAPTARF-----GDPEGFARF--VNGAHKVGIGVLLDWPAHFPTDEHG 360  
Qy 364 -----RKNKSAVVG-----DVLSSL-----EKAYKEGAK--AN 390  
Db 361 LWFPGTALYEHADPRQGFHPDWTATINFGRIEYMSVLNNALYWAERFLHDGLURDVA 420  
Qy 391 REMTYLSETL-----FGG---TEVVRFAQFANALATNPDAHAGILKSLDDKYDYL 438  
Db 421 ASMLYLSRKEGEWIPNEYGGRENLESVRFLOKNSLVYG--THPGVMTAEESTS--W 476

Qy 439 PSIDRKVLPAMLDIVRRRIIPADKLPIEFKNVIDKKFKGDTKKYADFVDFKSVVPSYDKFH 498  
Db 477 PKVSPQHEGGLG-----FGFKWNGFMHDTLSY-----FSRE--PVHRKFEH 516  
Qy 499 AM-----LKSMDKEKFAKIEKDPAVELSKSVIAAARAIQADAMANAAYAI-----EKGK 547  
Db 517 HQELTFGLLYAFTENFVLPPLSHDEVHVGKSLIAKMSDDWQKFAFNLRSYGGEMWGPYK 576  
Qy 548 RLFPAG-----LRE--MYPG--RALPDSANFTMRMSYG--SINGYEPQD 585  
Db 577 KLLFMGQFAQWSEWSEKSLDWNLRQYPMHGMRRRLVRODLNLTYSKAAALHARDCEPDG 636  
Qy 586 GAW--YNYHTTGKVLKQDPKPSDEFAVOENLILDFTKNY-----GRYAENQOLHIAF 637  
Db 637 FRWLVDVDDHENSFVAMLRTPAGKPVAVICNLTPVYRENYVYVLPVAVAGRWRE----- 688  
Qy 638 LSNND--ITGNSGSPVFDKNGRLIGLAFDGNWEAM-----SGDIEFEPD 680  
Db 689 ILNTDAEYGG--SGK---GNGGRVQAVDAGEIGACAMLVLPPLATIMLEPE 734  
RP SEQUENCE FROM N.A.  
ID Q9LQJ8 PRELIMINARY; PRT; 1808 AA.  
AC Q9LQJ8; 2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE F28G4.7 protein.  
GN F28G4.7  
OS *Arabidopsis thaliana* (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
RA Buehler E., Chao Q., Chiu C., Chiu J., Choi E., Gonzalez A.,  
RA Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
RA Lenz C., Liu A., Liu S., Mukharsky N., Pham P., Sakano H., Shinn P.,  
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.,  
RA Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AC007843; AAF97311.1;  
DR EMBL; AC007843; AAF97311.1;  
SQ SEQUENCE 1808 AA; 202730 MW; 9B805C7E2B533EB1 CRC64;

Query Match 3.3%; Score 124; DB 10; Length 1808;  
Best Local Similarity 18.7%; Pred. No. 60;  
Matches 157; Conservative 113; Mismatches 289; Indels 280; Gaps 38;

Qy 23 KADKGMILLNEL--NOENLD-----RMRELGETLP-----LDSLYS 56  
Db 643 KSNELIPLINDIQRQTNVDLDARASKSGAMRANGFVLAHVYKLLHCFGLDWFSSLS 702  
Qy 57 FDKPSIANAVIFGGCGTGITVSDQGLFTNHHGCGYGAISQSTVDHDPYLRDGFVSRMTG 116  
Db 703 WDN-----AFSIHDKQSDNLFALDFAKAMPLE 731  
Qy 117 EELPIPLGLS-----VKYLRKIVKVD-----KVEGOLKGITDEMELRLKRAQ 158  
Db 732 LFLQVGVSTQKADMMKKCKQVMRLSELPGEEYKLLMDTLATGRLSMLDILRLKLQ 791  
Qy 159 VCOELAKKENADENQLCTVEPFYSNNEYFLIVYDFKDVRMVFAPPSSVGFPGGDTDNM 218  
Db 792 VSSRLRRDE-----IEEKYANLTH-----AMELKPYIEEPVFAATSVNM 831  
Qy 219 WPRHTGDSFVRVYAGADNRPA---EYKDNKPKYPVFAAVSMQGYKADYAMTIGPPG 275  
Db 832 -----SLDFRPRIRHDFILSRDAVDEWLTLEYCYAAADHRAAKLAEPG 876  
Qy 276 STDRYL---TSWGVEDRIENENNPRIEVR--GKQGIWKEAMS----- 313



Db	877	SVVOEVRFRS	WASDRVMTTEQ	RAKLLKRIAIDE	---	KEKLSFCECEKIAKDLNLTUEQV	933											
QY	314	-----ADQ	ATRIKASYKAQ	SANYKNSIGM	NRGLARDLVIG	KRAE--ERAFADWIRK--	365											
Db	934	MHYHAKHGR	RVKSKDKHLA	-	IDNSSSSSG	-----KKRGTLVKTGTGEVRSII	984											
QY	366	-NCKSAVYGDV	-----	LSSLEKAYKEGA	KANREMYLSE	TLFGTEVVRFAQFANA	415											
Db	985	VDEKVLNSDA	IDASNSEKFL	NLSLEEHOEHLN	LOENSEIRDTE	DEGQCSSIIN--QYASS	1042											
QY	416	LATIN-PDAH	AGILKSLDDK	---YKDYLP	SLDRKV-----	LPAMLDIVRRIPAD	460											
Db	1043	KYTSTPSQ	RSQFWTDEADR	KLLSQYVHR	RAALCAKEFH	GVWVASVDELPA	PLPACARRVOIL	1102										
QY	461	KLPDIFKN	VIDKKFGDT	KYKADYFVDK	-SVVPYS	SDKFHMLKSM-----	504											
Db	1103	MKNDKPKA	IMSLCNLLS	RYARHLETK	OKCLPESN	SHVLVRVLS	SPAIGTDSGSVEQG	1162										
QY	505	-----DKEK	FAKAIK-----	DPAVELSK	-SVTAAARA	TOADAMANA	YAIKGRLLFF	5515										
Db	1163	KOICFDEE	KWDNFNEKS	ISQAFNDVLEL	KMAKLVP	KRTKSSREWS	NRDIIDEGS	1218										
QY	552	AGLREMYG	RALPSDANET	MRMSYSGIK	EYEPQDCAW	NYHTTGK	VLEKODPKSD	DEFV	611									
Db	1219	---EMV	PAIHSED---	IQNVSV	DOVKDTS	SPRSG-HY	LELHQTVP	PLDEKD	---NDSIQV	1267								
QY	612	QE-----	-----	NIL-DLF	FTKNYGRYA	ENGQHLIA	FLSNND---	642										
Db	1268	RKSLAV	STAELLKLV	LSMPTAG	MPNLLDTLR	-----	RYSER-DL	FATAYSL	LRKKF	1321								
QY	643	ITGNGSGSP	-VF	DKN-----	-----	GRLGIA	FDGNWEANS	GDI	EPE	DLQ	682							
Db	1322	LVGGSGGQ	FPVLSQ	NFLHSIS	KSPVNTG	TAAKFS	WLSFEL	HERDL	MAGG	VTIT	SDLQ	1380						
RESULT	10																	
Db	085051																	
ID	085051																	
AC	085051;																	
DT	01-NOV-1998	(T	REMBL	rel. 08, Created)														
DT	01-NOV-1998	(T	REMBL	rel. 08, Last sequence update)														
DT	01-JUN-2001	(T	REMBL	rel. 17, Last annotation update)														
DE	Transferrin	binding	protein	B.														
GN	TBPB.																	
OS	Moraxella	catarrhalis.																
OC	Bacteria; Proteobacteria;																	
OX	NCBI_TaxID=480;																	
ON	[1]																	
RP	SEQUENCE FROM N.A.																	
RC	STRAIN=LES-1;																	
RX	MEDLINE=98380363;																	
RA	Myers L.E., Yang I.P., Du R.P., Wang Q., Harkness R.E.,																	
RA	Schryvers A.B., Klein M.H., Loomsore S.M.;																	
RT	"The transferrin binding protein B of moraxella catarrhalis elicits																	
RT	bactericidal antibodies and is a potential vaccine antigen."																	
RL	Infect. Immun. 66:4183-4192(1998).																	
DR	EMBL: AF039313; RAC34278.1; .																	
DR	InterPro: IPR001677; Transferrin_bind.																	
DR	Pfam: PF01298; Lipoprotein_5; 1.																	
SQ	SEQUENCE	713	AA;	76811	MW;	A9244	5A61E	A905B	D	CRC64;								
Query	Match																	
Best	Local Similarity																	
Matches	123; Conservative																	
QY	119	LP	IPGLSV	RYLRKIV	KVYTDK	VEGQLK	IGITD	EMERL	URKAQ	EVQC	QELAK	NADENQ	LCTIVE	178				
Db	156	IP	DPK	NLIEV	LKSS	SVSKS	FEAQ	GGIENN	TRLTK	DLK	LS	SEQ	EAKY	KEALDNAL	----	211		
QY	179	P	FYSN	NEY	ELIV	---YD	PEK	VRV	MFAP	SPSS	VKG	EGG	DTD	NM	MPRHTG	DFSV	RVYAGA	235
Db	212	TO	PAQ	E	KY	RELI	ENIA	HA	DKK	SDAR	-----							234

[illegible]

Query Match 3.3%; Score 123; DB 16; Length 1289;  
Best Local Similarity 18.1%; Pred. No. 40;  
Matches 125; Conservative 111; Mismatches 256; Indels 200; Caps 29;

```
QY 102 DH DY LR DG FV SR TM GE EL PI PGL SV KY LR KI VK VT DK VEG QL KG ITD----- 148
```

Db 22 DSDALR--YRLNTISE-----LNKSYLKAIONIVG-YNGLFKSGSDFFLFLETYSNDPLL 72  
Qy 149 -----EMERLRKAQEVCOELAKKENADENOLCIVEPEYSNNEYFL----- 188  
Db 73 ARYOKESQEAIRKVOEARREGASNEEILKLQLOAEARLRNKKQFIALIESNLNSIDYL 132  
Qy 189 -----IVVDVFKDVRWVFAPPSSVGKFGGDTDNWMPRHGTDFSV-PRVYA 233  
Db 133 ASETLKNSLKTANKIIDDVLKNWKEILKPLE-GDFGVKINAILQLIDSTYKITTIDYC 191  
Qy 234 ---GADNRPAEYKDNKPKYPVFAAVSMQYKADYAMTIGPGSTDRYLT--SWGVE 287  
Db 192 KANYGATEKIAQY-----WWTYF-----VEPFESSEKKVL-----DEVLTKASDELK 235  
Qy 288 DRIENENNPRIEVRGIKQ-----INKEA--MSADOATR-----IKYASKYAQSA 330  
Db 236 DRLKDLRLTERLKLIVSKGNALAEKVTEETAKKSAESILKNIETPSLIVELFTKYNNV 295  
Qy 331 NY--WKSIGNRGL---ARLDVIGRKRAE-----ERAFADWIRKNGKSAYGDIVLSSL 379  
Db 296 DFOITFNDIASNEVVFQIRELVGNDENEINRCYFDRAYFLSKKAAKAAIINDVGEKT 355  
Qy 380 EKAYKEGAKANREMYLSETLFGTEVVRFAQAFANALATNPDAHAGILKSLDDK----- 433  
Db 356 QGSGTKGORKSEEDVIOEVIAKAEIV-----LAKDFLKVSAVQNTETRIEEMND 407  
Qy 434 --YKDYLPSLDR-----KVLPAMLDIVRRRIPADK 461  
Db 408 LSLRDALPDLSAKRLIETFTGOLENNEITLGEYNEINRIVSVYTGOLNTAKTTIANNL 467  
Qy 462 LPDIFKNVI-----DKFKGDTKKYADFVF-----DKSVVPYSKDFHAMLKS 503  
Db 468 ESEYHKGSIQSEYSEKQLDKDQVEGNIQOLNDFVVLKHOEIKQRLFKELSLQVKT 527  
Qy 504 MDKEKFAKAEKDPAVELSKSVIAAARAIQADAMANAIAIEKGRLF---FAGLREMYPG 560  
Db 528 LDQSEMKL-----LQDLSLA-----KFNQFLKLYKIGQAVSWFAFSESEKMEG 574  
Qy 561 RALPSDANFTMRSGYSGIKYEPQOGAWYNYHTTG-----KGVLKODPKSDFA-A 610  
Db 575 FTNGIGRAFMKELSKSPSPFLQNGGWIENLALICOEDERLTLALIEKVDKLSEIKA 634  
Qy 611 VQENILDLFRTKNYG--RYAENGOLHIAFLSN 640  
Db 635 LYESFPYCYTDEGLVSYSKNAMLIEEFKSS 666

## RESULT 12

Q92B47 Q92B47 PRELIMINARY; PRT; 742 AA.  
AC Q92B47  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Cpa.  
GN Cpa.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS101;  
RA Podbielski A., Woischnik M., Leonard B.A.B., Schmidt K.H.;  
RT "Characterization of nra, a global negative regulator gene in group A  
streptococci";  
RL Mol. Microbiol. 31:0-0(1999).  
DR EMBL; U49397; AAC97148.1; -  
SQ SEQUENCE 742 AA; 83683 MW; 97A1FF44B4ECB944 CRC64;

## Query Match

3.3%; Score 122.5; DB 2; Length 742;

Best Local Similarity 21.2%; Pred. No. 18;

Matches 133; Conservative 62; Mismatches 212; Indels 221; Caps 31;

Qy 206 SVKFGGDTDNWMPRHGTDFSVFVYAGADNRPAEYKDNKPKYPVFAAVSMQ---Y 262  
Db 40 SIRAFAGEQS--VPNRQSSITQDYPWY-GYDYPKGY-PDYSPLKTYHNLKNVLESGSKDY 95  
Qy 263 KADYAMTIGPGSTDRYLSWGVEDRIENEN-----NPRIEVRGIKQGIWKEAMSADQ 316  
Db 96 QAYCFNLTKHFPKSDSVRSOWYKKLEGTNENFIKLADKPRIEDGOLQONFI----- 146  
Qy 317 ATRIKYASKYAQSAANYKNSIGMNRGLARLDVI-----GRKAEERAFADWIRK 365  
Db 147 -LRILY-----NGYPNNRNGIMKGDPLNAILVTQNAIWTDSAQINPDESFKTEARS 198  
Qy 366 NGKSAVYCDVLSSLEKAYKEGAKANREMYLSETLFGTEVVRFAQAFANALATNPDAHAG 425  
Db 199 NG---INDQQLGLMRKALKELIDPNLGSYSNKTTPSG-----YRLN 236  
Qy 426 ILKSLDDKYK-----DYLPSLDRK---VLPAMD-----IVRRRIPAD---KL-- 462  
Db 237 VFESHDKPFQNLLSAEYVVDTPPKPGGEPPEAKTEKTSVIIRKYAEGSKLLEGATLKLQ 296  
Qy 463 -----PDIFKNVIDKKFKGDTKKYADFVFDK 488  
Db 297 IEGSGFQEKDFQSNLSGETVELPNGTYTLTETSSPDGTYKTAEPKFRVENKKV--FIVQK 354  
Qy 489 S-----VVVPSDKFHAMLKSMDE-----KFAKAEKDPAVELSKSVIA- 527  
Db 355 DGSQVENPNKEVAEPYS--VEAYNDFMDEEVLSGFTPYGKFFYATNKDKSSOVVYCFNAD 412  
Qy 528 -----AARATQADAMAN---AYAIEKGRKLFPAFLREMYPGRALPSDAN---FTMR 572  
Db 413 LHSPDPDYDSGETINPDSTMTKVKYTHTAGSDILFKYALR-----PRDTPNPEFLKH 464  
Qy 573 MSYSGIKYEPQOGAWYNYHTTGKGVLEKODPKSDFAVO--ENILDLFRTKNYGRYAE 630  
Db 465 IKKVIKGYK-KKGDSYN-----GLTETOFRAATQALAIYFTDSADLTKLTKTY-----NN 513  
Qy 631 GOLHIAFLSNND-----ITGNSGS-----PVFDKNGRLIGLAFDGNWEA 670  
Db 514 GKGYHGESMDEKTLAVTKELITVAQNSAPOLTNLDFEVPNNSKDQSLIG----- 564  
Qy 671 MSGDIEFEPDLQRTISDIRYVLFMIDK 698  
Db 565 ---TECHPD-----DLVDVIRMEDK 581  
RESULT 13  
Q92B35 Q92B35 PRELIMINARY; PRT; 1571 AA.  
AC Q92B35  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein lin1716.  
GN Lin1716.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
BAguero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fsihi H., Garcia-Gel Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordisiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,

[illegible]

RESULT 14
Q99FL1
ID Q99FL1
AC Q99FL1

```
QY 539 NAYAIEK-----GKRLFPAGLREMYTPGRALP-SDANFTMRMSYSGSIKGYEPQD----- 585
Db 1065 DPPVPXPRACKKCGCCGKNHNYPDNLNPFSESGFWSRFKKGHFQG-PMQDVSSLINIL 1123
QY 586 -GANVYHTTGKGVLEKQDPKSDFAVOENILDLFRTKNYGRY 627
Db 1124 KGAEMIYHQPEKVIKWLKTWRTAEYVSSDFLOA-QMKDIPHY 1165

RESULT 15
Q92BR7
ID Q92BR7 PRELIMINARY; PRT; 721 AA.
AC Q92BR7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein lin1477.
GN Lin1477.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RA PubMed=11679669;
RX Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,
RA Maqueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596168; CAC96708.1; -.
DR ListList; LIN01477; -.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR001460; Transpeptidase.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 721 AA; 79910 MW; D0A0D8167AA6837E CRC64;

Query Match 3.38; Score 122; DB 16; Length 721;
Best Local Similarity 22.1%; Pred. No. 18;
Matches 112; Conservative 69; Mismatches 195; Indels 130; Gaps 24;

QY 277 TDRVLTSGVEDRIENENPRIEVRGKQG-----IWKEAMSADQATRI----- 320
Db 123 TDRDLKDYWILTHQTESLNLRLSAKEQALESSKAYKIQVEKVTNDDIAKLTTLTKVATIIY 182
QY 321 -KYASKYAQSANYNKNSGTMNRGLARLDVIGRKRAERAFADWIRKKNKSAVYGDVLSL 379
Db 183 KKMTGTGAMTESVIKNDVDEETARVENMDSLPGVDTTTDDNRYRYTETLRSILGSV 242
QY 380 EKAYKEGAKANREWTYLSSETLFGGTEVRF---AQFANALATNP-----DAHAGILKS 429
Db 243 STA-KEGLPKDKAEYILSQGYSDNRDRGVKSYLEAQYESVLGASQSKSSESVDLSKGNIIET 301
QY 430 LD----DRYKDYLPDLRKVLPAMLDIVRRIPADKLPDIEKNVIDKKFGDKTKYA-DF 484
Db 302 VSKYEGSGKGLVLSVDVFEQKAVEDILRNKIK-----QGKQYAGSD 343
QY 485 VFDKSV-----PYSDKFHAML--KSMOKEKFAKIEKDPAVELSKSVTAARAATQAD--- 535
Db 344 LFDRAFYVAMPDPYSGEVLALAGKLNDRKGEF-----DDYSLCTFTTAYAMGSAVKGSTIL 398
QY 536 -----AMAN-----AYAIKGRKLF-----AGLREMYPGRALPSDAN---FTMRM 573
```

```
Db 399 GGIMDGAISNKTVFTDQPIALKGTGPKSSWFNRTGAGNRPLDPVGALEISSNSYMYQVAM 458
QY 574 SYGSIKGYEPQ-----DGAWYNYHTTGKGVL-----EKQDPKSDFAVOENIL 616
Db 459 KMGAK-YVENGPLRAPLSTFDDMRYYNQFGLGVKTGIDLPGEQVGYKGDQDTIGK-IL 516
QY 617 DLFRTKNYGRYAENGQHLIAFLSNNDITGNSGSPVFDK-----NGRLIGLAFDGNWE 669
Db 517 DF-----AIGQYDSYTPLQMAQYVSTIANGSGRIAPSMVKETRNPFSTNGSDSVGTLATAN-- 570
QY 670 AMSGDIEFEPEPDLQRTISV---DIRYV 692
Db 571 -----EPKVLNKGVSNDIAKTV 588

Search completed: May 25, 2003, 15:36:06
Job time : 64.8284 secs
```

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2003, 15:31:51 ; Search time 14.1927 Seconds  
(without alignments)  
2080.729 Million cell updates/sec

Title: us-10-008-355-2

Perfect score: 3719

Sequence: 1 MOKLKSILLGALLLGASG.....LFMDKWGQCPRLIQELKLI 712

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123.5	3.3	776	1 YLA4_CAEEL	Q05036 caenorhabdi
2	122	3.3	1104	1 NAGH_CLOPE	P43153 clostridium
3	121	3.3	1628	1 NAGH_CLOPE	P26831 clostridium
4	118	3.2	698	1 EFG_VIBCH	Q9kuz7 vibrio chol
5	116	3.1	952	1 TOPI_STRCO	Q9x909 streptomyce
6	114	3.1	620	1 DNAK_PORPU	P30723 porphyra pu
7	114	3.1	734	1 GLGB_AGRTU	P52979 agrobacteri
8	114	3.1	747	1 VIVC_BPT7	P03725 bacterioph
9	113	3.0	1047	1 RIRL_CHLMU	Q9p193 chlamydia m
10	112.5	3.0	657	1 CN16_HAEIN	P44764 haemophilus
11	112	3.0	556	1 EST2_CAEEL	Q07085 caenorhabdi
12	112	3.0	1504	1 DPO2_YEAST	P14284 saccharomyc
13	111.5	3.0	878	1 ACON_RICPR	Q9zcf4 rickettsia
14	111.5	3.0	922	1 YKL6_CAEEL	P42173 caenorhabdi
15	111.5	3.0	2410	1 MOK1_SCHPO	Q9usk8 schizosacch
16	110.5	3.0	1039	1 SY1_METJA	Q58357 methanococc
17	109.5	2.9	681	1 PARC_NETGO	P48374 neisseria g
18	109	2.9	768	1 P10_HUMAN	Q00566 homo sapien
19	109	2.9	3672	1 LML2_CAEEL	Q21313 caenorhabdi
20	108.5	2.9	507	1 YY42_CAEEL	Q18416 caenorhabdi
21	108	2.9	638	1 NEC2_PIG	Q03333 sus scrofa
22	108	2.9	1341	1 RPAL_METJA	Q58445 methanococc
23	107.5	2.9	666	1 UVRE_CLOAB	Q971q2 clostridium
24	106.5	2.9	747	1 Y030_UREPA	Q9prb5 ureaplasma
25	106	2.9	1592	1 GTF2_STRDO	P27470 streptococc
26	106	2.9	1813	1 UN13_CAEEL	P27715 caenorhabdi
27	105.5	2.8	663	1 UVRE_FUSUN	Q8grt2 fusobacteri
28	105	2.8	1023	1 HLY1_ECOLI	P09983 escherichia
29	105	2.8	1391	1 RPOB_MYCPN	P78013 mycoplasma
30	104	2.8	616	1 ACOC_SOLTU	O04916 solanum tub
31	104	2.8	908	1 H104_YEAST	P31539 saccharomyc
32	104	2.8	1060	1 DP3A_LACLA	Q9ci70 lactococcu
33	104	2.8	4563	1 APE_HUMAN	P04114 homo sapien

Query Match 3.3% Score 123.5; DB 1; Length 776;  
Best Local Similarity 21.1%; Pred. No. 2.5;

RESULT 1

ID	YLA4_CAEEL	STANDARD;	PRT;	776 AA.
AC	Q05036;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical 86.9 kDa protein C30C11.4 in chromosome III.			
GN	C30C11.4.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RC	MEDLINE=94150718; PubMed=7906398;			
RA	Willson R., Ainscough R., Anderson K., Baynes C., Berks M.,			
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,			
RA	Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,			
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,			
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,			
RA	Latrelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,			
RA	Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,			
RA	Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,			
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,			
RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,			
RT	Wooldman P.;			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.			
RT	elegans."			
RL	Nature 368:32-38(1994).			
CC	-1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,			
CC	TO YEAST MSI3.			

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL: L09634; AAA27967.1; --  
PIR: S44784; S44784.  
WormPep: C30C11.4; CE00103.  
DR InterPro: IPR001023; Hsp70.  
DR Pfam: PF00012; Hsp70; 1.  
DR ProDom: PD000089; Hsp70; 1.  
DR PROSITE: PS00297; Hsp70; 1; FALSE\_NEG.  
DR PROSITE: PS00329; Hsp70; 1.  
DR PROSITE: PS01036; Hsp70; 2; 1.  
KW Hypothetical protein; ATP-binding; Multigene family.  
SQ SEQUENCE 776 AA; 86896 MW; 20FA975FB649FA9C CRC64;

CC	preference is shown for Gly ac P3 and P1'; Pro and Ala at P2 and	
CC	P2'; and hydroxyproline, Ala or Arg at P3'.	
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M9.	
CC	-!- SIMILARITY: CONTAINS 1 PKD DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL'outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
EMBL:	D13791; BAA02941.1; -	
DR	EMBL; AP003185; BAB79879.1; -	
DR	EMBL; D50309; BAA08848.1; -	
DR	MEROPS; M09.002; -	
DR	InterPro; IPR002169; Micollptase.	
DR	InterPro; IPR000601; Zn_domain.	
DR	InterPro; IPR000130; Zn_Mtpeptdse.	
DR	Pfam; PF00801; PKD; 1.	
DR	Pfam; PF01752; Peptidase_M9; 1.	
DR	PRINTS; PR00931; MICOLLPTASE.	
DR	SMART; SK00089; PKD; 1.	
DR	PROSITE; PS00093; PKD; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.	
KW	Hydrolase; Metalloprotease; zinc; Zymogen; Signal; Complete proteome.	
FT	SIGNAL 1 36	POTENTIAL.
FT	PROPEP 40 86	
FT	CHAIN 87 1104	MICROBIAL COLLAGENASE.
FT	DOMAIN 774 862	PKD.
FT	METAL 502 503	ZINC (CATALYTIC (BY SIMILARITY)).
FT	ACT_SITE 503 503	BY SIMILARITY.
FT	METAL 506 506	ZINC (CATALYTIC (BY SIMILARITY)).
FT	CONFLICT 38 38	L -> F (IN REF. 1).
FT	CONFLICT 722 722	I -> M (IN REF. 1).
FT	CONFLICT 748 748	G -> E (IN REF. 1).
FT	CONFLICT 945 945	V -> E (IN REF. 1).
FT	CONFLICT 970 970	T -> A (IN REF. 1).
FT	CONFLICT 987 987	A -> E (IN REF. 1).
FT	CONFLICT 1098 1098	I -> T (IN REF. 1 AND 3).
SEQ	SEQUENCE 1104 AA; 125935 MW; F4B7377194ED021C CRC64;	
Query Match. 3.3%; Score 122; DB 1; Length 1104;		
Best Local Similarity 18.0%; Pred.No.5;		
Matches 141; Conservative 105; Mismatches 278; Indels 258; Gaps 36;		
QY	29 WLLNELNQENL-----DRMRELGETPLDLSYSDKPSIANAVVFGGCT---GITVSD 80	
DB	419 WASKEVAQGMRVVONDKALEGHPDDTLTVIIVNSPEEYKLNRLINGFSTDNGGIYEN 478	
QY	81 QGLLFTNHHGCGYAQS-QSTVDHD---YLRDGFVSRTMGELPIGVL---SVKYLKRV 133	
DB	479 IGTFEFTVTPPEESLYTLEELFRHEFTHYLOGRYV-----VPGMWGGEFFQEGVL 529	
QY	134 KVTDKVEQLKGITDENELRKAQEVQELAKKENADENQICIVEPFYSNNFYFLIVDV 193	
DB	530 TWYEEGTAEFFAGSTRDTGDKPRKSVTQGLAYDRNRMSLYGVLHAKYGSMDFY----- 583	
QY	194 FKDVRMVFAPSSVGKFGGDDTNMWRPHRTGDFSFRVYAGADRNPAPYSKDNKPKPVY 253	
DB	584 -----NYGFALSNMYNNMGMF-----NKMWTYIKNN----- 611	
QY	254 FAAVSMOCYK-----ADDYAMTIGFGPGSTDRYLTSGWVEDRIENENNPRIE---VRGITK 304	
DB	612 ---DVSGYKDYIASMSSDYLGNKYQDYMDSLL-----NNIDNLDPVLSDEVYNG-- 659	
QY	305 QGIWKAEASDAQTR-IKYASKYAQ-SANYWKNSTGMNRGLARLDVIGRKAERAFADW 362	
DB	660 ---HEAKDINEITNDIKEVSNIKDLSNVEKSSQPFETYDMRGTVVGRSQGEEN---DW 712	
QY	363 IRKNGKSAVYGDVLSLSEK-----AYKEGAKANREMTYLSSETLFGGTVVVRFAQF 412	

RESULT 2

COLA\_CLOPE STANDARD; PRT; 1104 AA.

ID COLA\_CLOPE

AC F43153;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Microbial collagenase precursor (EC 3.4.24.3) (120 kDa collagenase).

GN COLA OR CP01073.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

ON NCBI\_TaxID=1502;

OX [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 87-113.

RC STRAIN=NCIB 10662;

RC MEDLINE=94110220; PubMed=8282691;

RA Matsushita O., Yoshihara K., Katayama S.-I., Minami J., Okabe A.;

RT "Purification and characterization of Clostridium perfringens 120-

RT kilodalton collagenase and nucleotide sequence of the corresponding

RT gene.";

RL J. Bacteriol. 176:149-156(1994).

RN [2]

RP SEQUENCE FROM N.A.

RP STRAIN=13 / Type A;

RC PubMed=11792842;

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;

RT "Complete genome sequence of Clostridium perfringens, an anaerobic

RT flesh-eater.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

RN [3]

RP SEQUENCE OF 1073-1104 FROM N.A.

RC STRAIN=NCIB 10662;

RA Matsushita O.;

CC CC -1- CATALYTIC ACTIVITY: Digestion of native collagen in the triple

CC helical region at xaa-1-gly bonds. With synthetic peptides, a

CC

FT CONFLICT 722 722 I -> M (IN REF. 1).

FT CONFLICT 748 748 G -> E (IN REF. 1).

FT CONFLICT 945 945 V -> E (IN REF. 1).

FT CONFLICT 970 970 T -> A (IN REF. 1).

FT CONFLICT 987 987 A -> E (IN REF. 1).

FT CONFLICT 1098 1098 I -> T (IN REF. 1 AND 3).

SQ SEQUENCE 1104 AA: 125935 MW: F4B7377194ED021C CRC64;

Query Match 3.3%; Score 122; DB 1; Length 1104;

Best Local Similarity 18.0%; Pred No. 5;

Matches 141; Conservative 105; Mismatches 278; Indels 258; Gaps 36;

QY 29 WLLNELNQENL-----DRMRELGLTLPDLSLXSFQDKPSTANAVVIFGGCT---GITVSD 80

DB 419 WASKEVKAQEMRVVQNDKALEEGNPDDILTVVIYNSPEEYKLNRIINGFSTDGGIYIEN 478

QY 81 QGLIFTNHHCGYGAIOG-QSTVDHD---YLROGFVSRVTGWEELPTPLG---SVKYLRLKIV 133

DB 479 IGTFIVERTPEESIVTLEELFRHEFTHYLOGRYV-----VPGMQGQEFQOEVGL 529

QY 134 KVTDDKEGQLKGTIDEMERLKRKAQEVQCQELAKKENADENQLCTIVEPFYSNNFYFLIVDV 193

DB 530 TWEEGTAEFFAGSTRDGIKPKRSVTOGLAYDRNRMISLYGLVHAKYCSWDF----- 583

QY 194 FKDVVMVFAPPSSVGRKFGGDTDNMMPRHTGDFSVFVYAGADNRPAEYSKONKPKPVY 253

DB 584 -----NYGFALSNYMYNNMMGMF-----NKMNTYIKN----- 611

QY 254 FAAVSMOGYK-----ADDYAMTIGFPGSTDRLYLSWGVEDRIENENNPRIE---VRGIK 304

DB 612 ----DVSQYKDYTASMSDYGCLNDKYQDYMDSL-----NNIDNLDVPLVDEYVNG--- 659

QY 305 QGIWKEAMSADQATR-IKYASKYAO-SANYWKNSIGMNRGLARLDVIGRKRAEERAFADW 362

DB 660 ----HEAKDINEITNDIKEYSNIKDLSSNVEKSOQFTTYDMRGTYVGVRSQGEEN---DW 712

QY 363 IRKNGKSAVYGDVLSSLEK-----AYKEGAKANREMTYLSLETGGTEVVRFAQF 412

Db 713 KMNSK---LNDILKELSKSWNGYKTVTAYFVNHKVDGNGNVYDVVFGHMT----- 763  
QY 413 ANALATNPDAHA-----GILKSLDDKYKDYLPGLDRKVLPAMLDIYRRRTIPADKLPIFK 467  
Db 764 ---DNTDTHVNKEPKAVIKS-----DSSV-----IVEEIN-----FD 794  
QY 468 NVTDKFKGDTKY-----ADFVDSKVVPSYDKFHAMLKSM-----K 506  
Db 795 GTEKSDGEGIKAYEWDGDKSNEAKATPKYNK-----TGEYEVKLTVDNNGGINT 849  
QY 507 EKFAKATEKDP-----AVELSKSVIAAARAQADAMANAIAEKGR--- 548  
Db 850 SKKIVVEDKPVINESEPNDFERANQAKSNMLVGLTSEEDYSDKYTFDVAKNV 909  
QY 549 -----LFFAGLREMY-----PCRALPSDANFTMRMSYSGIKGYEPQDG 586  
Db 910 KITLNNLSVGITWTLYKEGDLNNYLYATGNDGTVLKGEKTLPGRYLYSVTYDNQSG 969  
QY 587 AWNYHTTGKGVLEKQDPKSDFAVQENILDLFRKYNKYAENGQHLIAFLSNNDITGG 646  
Db 970 ---TYTVNVKGNLKEVKEAKDAIKE-----VEN-----NND----- 999  
QY 647 NSGSPVFDK-----NGRLIG-LAFDGNWEAMSGDTEFEPDLORTI-----SVDIRYVLF 695  
Db 1000 -----FDKAMKVDSNSKIYVGTLSNDDLKDIYSIDIONPSDLNIVVENLDNIKNWLLYS 1053  
QY 696 ID 697  
Db 1054 AD 1055  
RESULT 3  
ID NAGH\_CLOPE STANDARD; PRT; 1628 AA.  
AC P26831;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hyaluronoglucosaminidase precursor (EC 3.2.1.35) (Hyaluronidase)  
DE (Mu toxin).  
GN NAGH OR CPE0191.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN=CPN50;  
RX MEDLINE=94232189; PubMed=8177218;  
RA Canard B., Garnier T., Saint-Joanis B., Cole S.T.;  
RT "Molecular genetic analysis of the nagH gene encoding a hyaluronidase  
of Clostridium perfringens.";  
RL Mol. Gen. Genet. 243:215-224(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / Type A;  
RX PubMed=11792842;  
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
flesh-eater.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
CC -I- FUNCTION: Putative virulence factor which is likely to act on  
connective tissue during gas gangrene.  
CC -I- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-linkages between N-  
acetyl-beta-D-glucosamine and D-glucuronate residues in  
hyaluronate.  
CC -I- SUBCELLULAR LOCATION: Secreted.  
CC -I- CAUTION: The partially purified protein from strain CPN50 is  
approximately 70 kDa smaller than the sequence indicated here.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; M81878; AAA3259.1; -.  
DR EMBL; AF003185; BAB79897.1; -.  
DR PIR; S27540; S27540.  
DR InterPro; IPR002105; Dockerin\_1.  
DR InterPro; IPR000431; FA58\_C.  
DR Pfam; PF00404; Dockerin\_1; 2.  
DR Pfam; PF00754; FS\_P8\_type\_C; 1.  
DR SMART; SM00231; FA58C; 1.  
KW Hydrolase; Glycosidase; Toxin; Repeat; Signal; Complete proteome.  
FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 1628 HYALURONOGALACTOSAMINIDASE.  
FT DOMAIN 703 1032 3 X APPROXIMATE REPEATS.  
FT REPEAT 703 748 1.  
FT REPEAT 848 894 2.  
FT REPEAT 987 1032 3.  
FT VARIANT 147 147 G -> A (IN STRAIN CPN50).  
FT VARIANT 172 175 KIOS -> ELKN (IN STRAIN CPN50).  
FT VARIANT 250 250 V -> M (IN STRAIN CPN50).  
FT VARIANT 548 548 A -> E (IN STRAIN CPN50).  
FT VARIANT 558 558 D -> E (IN STRAIN CPN50).  
FT VARIANT 614 614 G -> S (IN STRAIN CPN50).  
FT VARIANT 944 944 I -> V (IN STRAIN CPN50).  
FT VARIANT 950 950 N -> S (IN STRAIN CPN50).  
FT VARIANT 979 979 T -> I (IN STRAIN CPN50).  
FT VARIANT 982 982 I -> L (IN STRAIN CPN50).  
FT VARIANT 1042 1042 I -> F (IN STRAIN CPN50).  
FT VARIANT 1043 1628 MISSING (IN STRAIN CPN50).  
SQ SEQUENCE 1628 AA; 182474 MW; D4252A2512BBED69 CRC64;  
Query Match 3.3%; Score 121; DB 1; Length 1628;  
Best Local Similarity 20.1%; Pred..No. 9.8;  
Matches 129; Conservative 78; Mismatches 209; Indels 226; Gaps 36;  
QY 176 IVEPFYSNN---EYFLIVYDVFKDVMFVAPPSSVSGKFGDGTNDNMWPHRTGDFSVFRV 231  
Db 111 VVDNYFNKNPHDESF---FDEKMDANIVSKVDCVIGVIGEDTDSAFY---GYTTLKHV 163  
QY 232 YAGADNRPAYSKDNKPKVPYFAAVSMQGYKADYAMTIGFPGSTDRYL-TSGMVEDRI 290  
Db 164 F---NQLEEGNK-----IQSFRADYYA-EVAHRGFIEGYGNPWSNEDRA 204  
QY 291 E-----NENNPRIEVRGKIQGIWEKMSADOATRIKYASKYASANYWK 334  
Db 205 ELMKFGGDYKLNQYVFAPKDDP-----YHNSKWRDLYPEKLESEIK---KLAQVGNETK 255  
QY 335 NSI-----GNRGLA-----RLDVIGRK-----RAEERAF----- 360  
Db 256 NRYVYALHPFNNPNVREDTEENYONDGLVIAKAKFTOLLENDVROFALLADDASAPAQAS 315  
QY 361 -----DWIRKNGKSAVYGDVLSL-----EKAYKECAK-----NREMYLSETLF 401  
Db 316 MYVKLLTDLTRWLEE--QOSTYPLDKTDLMFPCSDYDNGSSAQKLKLNKAEDNVSVMT 373  
QY 402 GGT---EVRFAQAFANALATN--PDHAG-----ILKSLDDKYKDYL-----PS 440  
Db 374 GGRIWGEVDE--NFANNFMANNISTEGHPGRAPFFWIKNPPCSNDSKQHLIMGNDTFLHPG 431  
QY 441 LDRKVLPAMLDIVRRRIPADK-----LPDIFKNVIDKFKGDTK-----KYADF---VF 486  
Db 432 VDPKIDGIVLNPQQAEANKSAFALADYAWNIWMDKEADENWNSFRYMDHGTAEET 491  
QY 487 DKSVPVYSDKFAHMLKSM--KEKFAKAEKDPAVELSKSVIAAARAATQADAM---ANAY 541  
Db 492 NSSLALREISKHMINQNMDGVRPLQESVELAPKLEAFKQKYDSCASIKEDALEIAEFT 551  
QY 542 AIEGKRLFTAGLRMPGRALPSDA-----NFTWRMSYSGIKG-----YEPQDGAW 588

```
Db 552 NLOKA-----ADYKKNPGRNTRDQIIYWLNCWEDTMDAAIGYLSAIAIEEGDDEAAW 606
QY 589 YNHTTGTGKVGLEKDPKSDFAVOENILDFRTKNYG-----RYAENGOLHI----- 635
Db 607 ANY-SEAQAFAEK-----SKTYGFHYVDHTEYAEGVQHIVPFIKS 646
QY 636 -----AFLSN-NDITGGNGSGPVFDKN 656
Db 647 MGQNLVYVIGSIVDPNRIIATYISNRQDAPTGNPDN-IFDNN 687

RESULT 4
EFG_VIBCH
ID EFG_VIBCH STANDARD; PRT; 698 AA.
AC Q9KU27;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elongation factor G (EF-G).
GN FUSA OR VC0361.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=E1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Daddelberg J.F., Eisen J.A., Nelson W.C., Peterson J.D., Umayam L.A.,
RA Gili S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT *DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.
RL Nature 406:477-483(2000).
CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC the nascent protein chain from the A-site to the P-site of the
CC ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This Swiss-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004124; AAF93534.1;
DR HSSP: P13551; 1ELO.
DR TIGR: VC0361;
DR InterPro: IPR004540; EF-G.
DR InterPro: IPR000640; EFG_C.
DR InterPro: IPR004161; EFTU_D2.
DR InterPro: IPR000795; EF_GTPbind.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00679; EFG_C; 1.
DR Pfam: PF00009; GTP_EFTU; 1.
DR Pfam: PF03144; GTP_EFTU_D2; 1.
DR PRINTS: PR00315; ELONGATNFT.
DR TIGRFAMS: TIGR00484; EF-G; 1.
DR TIGRFAMS: TIGR00231; small_GTP; 1.
DR PROSITE: PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
Complete proteome.
KW NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 88 92 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 698 AA; 76927 MW; F44FDA4DB8FC4ECC CRC64;
```

```
Query Match 3.2%; Score 118; DB 1; Length 698;
Best Local Similarity 18.4%; Pred. No. 4.8;
Matches 120; Conservative 83; Mismatches 202; Indels 246; Gaps 26;

QY 8 ILGAALLL-GASGVAKADKGMW-----LLNELNQENLDRMRGLFTPLDLSL 54
Db 105 VLGAVVVFCGTSCVEPQSETVMRQADKGVPRMVFYVNMKMDRAGADFLRVVG----- 156
QY 55 YSFDKPSIANAVI-----FGGCTGTV-----SDOGLIFT----- 86
Db 157 -QIKHRLGANPVPIQLNIGABEEFKGVIDLKKAIAINWNEADQGSFTYEIPADMLELA 215
QY 87 ---NHGCGYGAIOQSQTVDHYLRDGEVS-----RTMGELPIPLGSLVKYLRKIV 133
Db 216 QENRNHLVEAAEAASEELMEKYLEDGSELSEVEIKQALRQRTINNEIVLAACGSFAFKNGV 275
QY 134 K-VTDKV-----EGQLKGTDEMERLKAQEOVCOELAKKENADENQICIVEPYPSN 183
Db 276 QAVLDVAVIEFLPSPDTPVPAIKIDD-----RENSVERHADDN-----EPF--$ 316
QY 184 NEYFLIVYDVFVKDVRMVFAPSPSSVGKFGGDTDNMMWPRHTGDFSVFVRYAGADNRPAEYS 243
Db 317 SLAFKIAIDPF-----VGSILFIRYSGVNVSGDAVY 348
QY 244 KDNKPKYKPVYFAAVSMQYKAD-----DYAMTIG----- 272
Db 349 NSVKQKKEREGRIVQMHANKRDEIKETRAGDIAAIGLKDVTTGDTLCDPNHVVILERME 408
QY 273 FPSTDRYLSMGVEDRIENENPNRIEVRGKQGIWKEAMSA-DOATRIKYASKYAGS-- 329
Db 409 FP-----EPVIQIAVEPRSKADQEKMGIALGKLAADPPSRVETDGTOTLI 456
QY 330 -----ANYKNSIGMNRGLARLDVIGRK--RAERAFADWIRKNGKSAVYG-- 373
Db 457 SGMGELHLDIIVDRMKREFGVCNVGKPPQVAYRETIKSKSEVEGKFRQSGRGQYGHVW 516
QY 374 -----DVLSSLEKAYREGAKANREMTYLSFTLFGGT 404
Db 517 LKTEPAEPGGGFVFDVDAIGGVIPKEFINPVAKGIEQMMNNGVLGYPVLDVKATLPDGS 576
QY 405 -EVVRFQAFANALATNPDAHAGILKSL-----DDKYKDYLPDLSDRK--VL 446
Db 577 FHDVDSSEAFKTAGSMFAFKGALGAQAQVLEPLMKVEITTPEDWMDVGVGLNRRRGII 636
QY 447 -----PAMLDIVRRRTIPADKLPDIFKNWIDKK-----FKGDTKKYAD 483
Db 637 EGMDEGPAGLIKLIHAKVP---LSEMFYATDLKSATOGRASYSNEFAEYAD 684
```

## RESULT 5

```
TOPI_STRCO
ID TOPI_STRCO STANDARD; PRT; 952 AA.
AC Q9X909;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Unwinding enzyme) (Swivelase).
GN TOPA OR SCO3543 OR SCH5.06C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
SEQUENCE FROM N.A.
```

```
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
```



RA Rabbinitz E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RT Hopwood D.A.:  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)." (2002).  
RL Nature 417:141-147(2002).  
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE  
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.  
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded  
CC DNA, followed by passage and rejoining.  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA  
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN  
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS  
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.  
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL035636; CAB38480.1; -;  
DR HSSP; P06612; 1ECL.  
DR InterPro; IPR002936; DNAPrim\_toprim.  
DR InterPro; IPR003601; DNATopI\_ATP\_bind.  
DR InterPro; IPR003602; DNATopI\_DNA\_bind.  
DR InterPro; IPR000380; Prok\_topoisomerase.  
DR Pfam; PF01131; Topoisom\_bac; 1.  
DR Pfam; PF01751; Toprim; 1.  
DR PRINTS; PR00417; PRTPLSMRASEI.  
DR SMART; SM00437; TOP1AC; 1.  
DR SMART; SM00436; TOP1BC; 1.  
DR SMART; SM00493; TOPRIM; 1.  
DR TIGRfams; TIGR01051; topA\_bact; 1.  
DR PROSITE; PS00396; TOPOISOMERASE\_I\_PROK; 1.  
DR Isomerase; Topoisomerase; DNA-binding; Complete proteome.  
KW ACT\_SITE 334 334  
FT ACT\_SITE 334 334 DNA CLEAVAGE (BY SIMILARITY).  
SQ SEQUENCE 952 AA; 103584 MW; 1C98DABD7A629295 CRC64;  
  
Query Match 3.1%; Score 116; DB 1; Length 952;  
Best Local Similarity 18.8%; Pred. No. 9.8;  
Matches 124; Conservative 96; Mismatches 279; Indels 162; Gaps 29;  
  
QY 105 YLRDGF-VSRMTG--EELPIPLSV--KYLRKIVKVTDKVEGQLKI-----TDMERLRK 155  
DB 105 YLRDGF-VSRMTG--EELPIPLSV--KYLRKIVKVTDKVEGQLKI-----TDMERLRK 155  
QY 29 YLPGYVVEASVGHIRDLPSGAEEVPEKVTGEVRRRLGVDVDEHPQIYVNVNADKKSQVK 88  
DB 29 YLPGYVVEASVGHIRDLPSGAEEVPEKVTGEVRRRLGVDVDEHPQIYVNVNADKKSQVK 88  
QY 156 AOEVCQE-----LAKKENADENQLC-----IIVEPFYNNVEFLIVYDFKD-VRMVFAPP 204  
DB 156 AOEVCQE-----LAKKENADENQLC-----IIVEPFYNNVEFLIVYDFKD-VRMVFAPP 204  
QY 89 LKLDKESDELFLATDEDEGEALHWLQELKPKIPVKR--MVFETDKDAIRAANVP 146  
DB 89 LKLDKESDELFLATDEDEGEALHWLQELKPKIPVKR--MVFETDKDAIRAANVP 146  
QY 205 SSVGK-----FGGDTDNMMPRHTGDFSVFRVYACADNRPAEYKDNKP 249  
DB 205 SSVGK-----FGGDTDNMMPRHTGDFSVFRVYACADNRPAEYKDNKP 249  
QY 147 RELNKLVDQAETRIILDRLYGYEVSPVLMKKVMPRLSAGRVQSVATFLVVERERIRAP 206  
DB 147 RELNKLVDQAETRIILDRLYGYEVSPVLMKKVMPRLSAGRVQSVATFLVVERERIRAP 206  
QY 250 KPVYFAVSMQGYKADYAMTIGFPGSTDRYLTSGWVEDRIENENNRIEVRIGKIQGIWK 309  
DB 250 KPVYFAVSMQGYKADYAMTIGFPGSTDRYLTSGWVEDRIENENNRIEVRIGKIQGIWK 309  
QY 207 RSAEY--WDLTGTATGRAGDASDPSSLVARLQT-----VDGRRVAQG--- 247  
DB 207 RSAEY--WDLTGTATGRAGDASDPSSLVARLQT-----VDGRRVAQG--- 247  
QY 310 EAMSADQATRIKYASK-YAQSANYWKNKSGNRLGLARLDVIGKRAERAPADW----- 362  
DB 310 EAMSADQATRIKYASK-YAQSANYWKNKSGNRLGLARLDVIGKRAERAPADW----- 362  
QY 248 --RDFSLGQLKSANTLHLDEANALAAALALENTRFVRSVESKPYRRSPYAPERTTLQ 305  
DB 248 --RDFSLGQLKSANTLHLDEANALAAALALENTRFVRSVESKPYRRSPYAPERTTLQ 305  
QY 363 ---IRKNGKSAVYGDVLSLSLEKAYKEG--AKANREMYTLSETLFGGTEV-VRFQAQANAL 416  
DB 363 ---IRKNGKSAVYGDVLSLSLEKAYKEG--AKANREMYTLSETLFGGTEV-VRFQAQANAL 416  
QY 306 QEASRKLGFCA--KSTMQVAQKLYENGYIYMRDSTLTSDTAVSARAQVTLGYADYL 363  
DB 306 QEASRKLGFCA--KSTMQVAQKLYENGYIYMRDSTLTSDTAVSARAQVTLGYADYL 363

## RESULT 6

## DNAAK\_PORPU

ID DNAAK\_PORPU STANDARD; PRT; 620 AA.  
AC P30723;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Chaperone protein duak (Heat shock protein 70) (Heat shock 70 kDa  
DE protein) (HSP70).  
GN DNAAK.  
OS Porphyra purpurea, and  
OS Porphyra umbilicalis (laver).  
OC Chloroplast.  
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.  
OX NCBI\_TaxID=2787, 2786;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P. purpurea; STRAIN=Avonport;  
RA Reith M.E., Munholland J.;  
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast  
RT genome".  
RL Plant Mol. Biol. Rep. 13:333-335(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=P. umbilicalis; STRAIN=Avonport;  
RX MEDLINE=92077124; PubMed=1720741;  
RA Reith M., Munholland J.;  
RT "An hsp70 homolog is encoded on the plastid genome of the red alga,  
RT Porphyra umbilicalis".  
RL FEBS Lett. 294:116-120(1991).  
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U38804; AAC08201.1; -;  
DR EMBL; X62240; CAA44160.1; -;  
DR PIR; S19660; S19660.  
DR HSSP; P04475; 1DG4.  
DR InterPro; IPR001023; Hsp70.  
DR Pfam; PF00012; HSP70; 1.  
DR ProDom; PD000089; Hsp70; 1.

[illegible]

```
Db 521 FGLLYAFTENFVLPLSHDEVHKGSLIAKMSGDDMQKFAANLRSYGFWMGYPGKKLLFM 580
Qy 553 G-----LRE--MYPG-RALPSDANFTMRMSYC-SIKGYEPODGAW-- 588
Db 581 GQEPAAQNSEWSEKSLDNWNLROYPHMGEMRRLVRDLNLTYSKAALHARDCPEPDGRWL 640
Qy 589 YNYHTTGKVLKQDPKDFEFAVQENILDLFRKNY-----GRYAENGQLHIAFLSNND 642
Db 641 VDDHENSFVAMLTAPGEKPVAVICNLTPVYRENYVPLGVAGRWRE-----ILNTD 692
Qy 643 --ITGNSGSPFDKNGRLIGLAFDGNWEAM-----SGDIEFEPD 680
Db 693 AEIYGG-SGK---GNNGRQVQVADAGGETGAMLVPLPLATIMLEPE 733

RESULT 8
VIVC_BPT7
ID VIVC_BPT7 STANDARD; PRT; 747 AA.
AC P03725;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DE Internal virion protein C.
GN 15.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10760;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RA Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements.";
RL J. Mol. Biol. 166:477-535(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01146; CAA24433.1; -
DR PIR; A04351; HIBPC7.
DR PIR; SA2331; S42331.
SQ SEQUENCE 747 AA; 84341 MW; 959C572B7B42C2B8 CRC64;

Query Match 3.1%; Score 114; DB 1; Length 747;
Best Local Similarity 18.8%; Pred. No. 9.4;
Matches 114; Conservative 86; Mismatches 190; Indels 218; Gaps 27;

Qy 91 GYGATQSQSTVDHDLRGFVSRTMGEELPCLSVKYLKLVKVTDKVEGOLKGITDMM 150
Db 26 GYRAATQAEQPRSSLLD-----TIQ-----RFAKAGADMVYTAKEQARDLADE- 69
Qy 151 ERLRAQEOVCELA--KKNADENQOLCIVEPFYS-----NNEYFLIVYDVFKQVR 198
Db 70 ----RSNEILRTPEQRREALNNGTLLYQDDPYAMEALRVKVTGRNAAVLYDDVDYQKIK 125
Qy 199 M-VFAPPSSVGFGGDTDNMMVPHRTGTDFSVFRVYA---GADNRPAEYSKDNKPKPVYF 254
Db 126 EGVFRTREMEY-----RISRLOEGAKVYAEQGDIPEDVDYOR----- 165
Qy 255 AAVSQMGYKADYAMTIFPGSTDRYLTSWGVEDRIENNNPRIEVRGKIQ----- 305
Db 166 -----GFGNDITERNISLYGAHDNFLSQQAQKGAIMNS---RVELNGVLQDPDMLRRPD 216
Qy 306 -----GIWKEANSAD-QATRI-----KYASKYAQSANYKNSIGMNRGLARLDV 348
Db 217 SADPFKEYIDNGLVTGAIPSDAQATOLISQAFSDASSRAGGADF-----LMR 263
```

```
Qy 349 IGRKRAERAFADWIRKNGKSAVYGDVL-----SSLEKAYK---- 384
Db 264 VGDKK-----VTLNGATTYRELIGEOWNALMVTAQRSQFETDAKLINEQYRLKIN 314
Qy 395 -----EGAKA-----NREMTYLSSETLPGGTEVVVRFQAFANALATNPDA 422
Db 315 SALNOEDPRTAWEMLOGIKAELDKVQPDQMTPOREWLTSAQE-----QVQNMNAWTKA 369
Qy 423 HAGILKSLDDKYKDYLPSLDRKRVLPAMLDIV---RRRIPADKLPDIPKKNVIDKKFKGDT 478
Db 370 QA---KALDDSMK-----SMNK-----LDVIDKQFKRINGEWSTDFKMPVNTGEF 416
Qy 479 KKYADFVFDKSVVPYSDRKFHMLKSM-----KEKFAKAIKQPAV----- 519
Db 417 K-----HSDMVNVANKKLAIEDSMIDPDGAKDAMKLYLQADSKDGAFRTAIGTMVTD 469
Qy 520 ---ELSKSVIAARAIAQADAMANAIAIEKGKRLFFAGLREMYPGRALPSDANFTMRMSYG 576
Db 470 AGQWSAAVINGKLPERTPAMDALRRIRNADPQLIAAL-----YPDQAEFLTMDMH 521
Qy 577 STKGVEPQ 584
Db 522 DKQGIDPQ 529

RESULT 9
RIRL_CHLMU
ID RIRL_CHLMU STANDARD; PRT; 1047 AA.
AC Q9PLJ3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonucleoside-diphosphate reductase alpha chain (EC 1.17.4.1).
DE (Ribonucleotide reductase).
GN NR04 OR TC0214.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RA pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY
CC FOR DNA SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC thioresoxin + H(2)O = ribonucleoside diphosphate + reduced
CC thioresoxin.
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002288; AAF39086.1; -
DR HSP; P00452; 5RIR.
```

```
DR TIGR; TC0214; --
DR InterPro; IPR005144; ATP.
DR InterPro; IPR000788; Ribonucleo.red.
DR Pfam; PF00317; ribonuc_red.lg; 1.
DR Pfam; PF02867; ribonuc_red.lgc; 1.
DR Pfam; PF03477; ATP-cone; 3.
DR PRINTS; PR01183; RIBORDASEM1.
DR PROSITE; PS00089; RIBORED_LARGE; 1.
KW Oxidoreductase; DNA replication; Complete proteome.
FT ACT_SITE 458 458 BY SIMILARITY.
FT ACT_SITE 672 672 BY SIMILARITY.
FT ACT_SITE 687 687 BY SIMILARITY.
FT ACT_SITE 1043 1043 INTERACTS WITH THIOREDOXIN/THIOREDOXIN
FT SITE 1046 1046 INTERACTS WITH THIOREDOXIN/THIOREDOXIN
FT SITE 1046 1046 (BY SIMILARITY).
FT SITE 1046 1046 (BY SIMILARITY).
SQ SEQUENCE 1047 AA; 120035 MW; 9B3FFD9BFAF817AA CRC64;

Query Match 3.0%; Score 113; DB 1; Length 1047;
Best Local Similarity 18.1%; Pred. No. 17;
Matches 111; Conservative 106; Mismatches 196; Indels 200; Gaps 32:

Qy 112 SRTWCEELPIGLSVKYLKIVKTYDKVEGOLKGTIDEMERLKRKAQVCOELAKKENADE 171
Db 37 TRRIDDDHMLP-----EDLENSIRSLTHQV-----VKEVVKIT-----D 71

Qy 172 NQLCTVEPFYSNNEVFLIVDFKDVRFVAFPPSSVSGKFGGDTDNMMWPRHTGDFSVFRV 231
Db 72 GOVTVVERIQDWESQYLI-NGLQDV-;-----ARDVVVYR- 104

Qy 232 YAGADNRPAEYSKO-----NPKYPVYFAAVSMOGYKADD--YAMTIGFGPGST 277
Db 105 ----DDKRAHREKSWQSLSVIRRCGTTVFHFNPMKTSAALEKAFRATDRIEGTMPTDFVREE 160

Qy 278 DRYLTS---WGVEDRIENNPRIEVRGIKGIGKWEAMSAQATRIKYASKYAOANY-- 332
Db 161 VNALQKTVAEIEECSSQDS-RIDIEQIQDIVEQQLM-----VVGHVATAKNITL 210

Qy 333 WKNSTGMNRLGLARLDVIGKRAEERAFADWIRKNGKSAVYGDVLSLEKAYKEGAKANRE 392
Db 211 YREA-----RARVRDNRVEDQIVEEAPSETEFVLS-----KDGSL- 245

Qy 393 MYLSETLFGGTEVVRFAQANALATNPDAHAGILKSLDDKYKDYLPDL-DRKVLPLAMD 451
Db 246 -TYMT-----THSQLLARLARACSRFPETTTDAALLT-DMAFSNFYSIGIKESEVVLACIM 297

Qy 452 IVRRIRIPADKLPD-----IFKNVIDKKFKGDTKKYADFVDFKSVVVPYSDKQFHA 499
Db 298 AARANI--EKEPDYAFVAAELLDDVVYKEALDRS-RGDED----- 334

Qy 500 MLKSMDEKFAKAIKEDPAVELSKSVIAAARAIOADAMANAYAIKGRKRLFPAGLRMY- 558
Db 335 -LEQVYRDHFKRY IEGDSYRLNPEL---KNLFDLDALANAMDLSRDLOFSYMGIONLYD 390

Qy 559 -----PGRALPSDANFTWRMSYG-SIKGYEPQDCAW-----YNYHTT-----GK 596
Db 391 RYFNHDDGRRLLETPQIFWMRVAMGLALK--EQDKTYWAIFFYNLLSTFRYTPATPLFNS 448

Qy 597 GYLEKQDPKSDFAVOENITLDFRTKNYGRYAENGOLHIAFLS-----NNDITG-GNSG 649
Db 449 GMRHSQLSYCYLSTVQDDLVNIYKV-----ISDN-----AMLSKWAGGICNDWTAIRATG 498

Qy 650 SPVFDKNRGLIGL 662
Db 499 ALIKGTNGKSGQV 511

RESULT 10
CN16_HAEIN
ID CN16_HAEIN STANDARD; PRT; 657 AA.
AC P44764;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update).
```

```
16-OCT-2001 (Rel. 40, Last annotation update)
2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (EC 3.1.4.16).
CDB OR HI0583.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD."
RL Science 269:496-512(1995).
CC -!- FUNCTION: THIS BIFUNCTIONAL ENZYME CATALYZES TWO CONSECUTIVE
CC REACTIONS CONVERTING 2',3'-CYCLIC-NUCLEOTIDE TO 3'-NUCLEOTIDE
CC AND THEN 3'-NUCLEOTIDE TO NUCLEIC ACID AND PHOSPHATE
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O =
CC nucleoside 3'-phosphate.
CC -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32740; AAC22242.1; -.
CC TIGR; HI0583; -.
DR InterPro; IPR002224; 5_nucleotidase.
DR InterPro; IPR004843; M_ppestrase.
DR InterPro; IPR004844; S/T_phosphatase.
DR Pfam; PF00149; Metallophos; 1.
DR Pfam; PF02872; 5_nucleotidase; 1.
DR PROSITE; PS00785; 5_NUCLEOTIDASE.1; 1.
DR PROSITE; PS00786; 5_NUCLEOTIDASE.2; 1.
KW Hydrolase; Multifunctional enzyme; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 657 2',3'-CYCLIC-NUCLEOTIDE 2'-
FT SIGNAL 27 657 PHOSPHODIESTERASE.
SQ SEQUENCE 657 AA; 72763 MW; 201CAAB415014499 CRC64;

Query Match 3.0%; Score 112.5; DB 1; Length 657;
Best Local Similarity 18.2%; Pred. No. 9.9;
Matches 122; Conservative 82; Mismatches 193; Indels 273; Gaps 36;

Qy 187 FLIVVDVDFKDVRFVAFPPSSVSGKFG-----GDTDNMMWPRHTGDF 226
Db 45 FLTDFDYK-----AP-----TDKFGFTRASLIRQARAEVKNSVLVDGDLTQGNPIADY 96

Qy 227 SVFRVYAGADNRPA-----EYSKDNKPKYPKYVFAAVSNGQYK 263
Db 97 QAAQGYKECKSNPAIDCLNAMNYEVGTLGNHEFNGLNVLAD--AIKQAKFPITVNSNVVK 154

Qy 264 ADDYAMTIGFGPGSTRYLTSWGVDR--LENNENPRIEVRGIKQG-----IWKEA 311
Db 155 A-----GTEEPYFTPYIQEKSVDNQG-----KTHKLKIGYGVFPQIMWMDKA 200

Qy 312 -MSAQDATR--IKYASKYQAQSNYKNSIGMNRGLARLDVIGKRAEERAFADWIRKNGK 368
```



RA Lemontt J.F., Lawrence C.W.;  
 RT "REV3, a Saccharomyces cerevisiae gene whose function is required for  
 RT induced mutagenesis, is predicted to encode a nonessential DNA  
 RT polymerase.";  
 RL J. Bacteriol. 171:5659-5667(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97103777; PubMed=8948103;  
 RA Purnelle B., Coster F., Goffeau A.;  
 RT "The sequence of 55 kb on the left arm of yeast chromosome XVI  
 RT identifies a small nuclear RNA, a new putative protein kinase and two  
 RT new putative regulators.";  
 RL Yeast 12:1483-1492(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97103777; PubMed=9169875;  
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,  
 RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
 RA Dietrich F.S., Dilius H., DiPaolo T., Dubois E., Duesterhoeft A.,  
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,  
 RA Hunnicke-Smith S., Hymen R., Johnston M., Kalman S., Kleine K.,  
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
 RA Marathe R., Messenguy F., Mewes H.-W., Mitiapati S., Moestl D.,  
 RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,  
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,  
 RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,  
 RA Walsh S.W., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,  
 RA Zhong W.W., Zollner A., Vo D.H., Hani J.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";  
 RL Nature 387:103-105(1997).  
 CC -1- FUNCTION: REQUIRED FOR INDUCED MUTAGENESIS, IT IS A NONESSENTIAL  
 CC DNA POLYMERASE. IT MAY FUNCTION IN TRANSLATION SYNTHESIS.  
 CC TRANSLATION SYNTHESIS IN S.CEREVISIAE MAY USE A SPECIALIZED DNA  
 CC POLYMERASE THAT IS NOT REQUIRED FOR OTHER DNA REPLICATIVE  
 CC PROCESSES.  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate  
 CC + [DNA] (N).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-B FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 CC EMBL; M29683; AAA34968.1; -;  
 CC EMBL; X96770; CAA65554.1; -;  
 CC EMBL; Z73523; CAA97873.1; -;  
 CC EMBL; A33602; A33602.  
 CC SGD; S0006088; REV3.  
 CC InterPro; IPR02064; DNA\_pol\_B.  
 CC Pfam; PF00136; DNA\_pol\_B; 1.  
 CC Pfam; PF03104; DNA\_pol\_B\_exo; 2.  
 CC PRINTS; PR00106; DNAPOLB.  
 CC SMART; SM00486; POLB; 1.  
 CC PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
 CC Transferase; DNA-directed DNA polymerase; DNA replication;  
 CC DNA-binding; DNA repair; Nuclear protein; Zinc-finger.  
 CC 2N.FING 1398 1417 C4-TYPE (POTENTIAL).  
 CC 2N.FING 1446 1473 C4-TYPE (POTENTIAL).  
 CC SQ SEQUENCE 1504 AA; 172956 MW; 542C6B664F734F5 CRC64;  
 Query Match 3.08; Score 112; DB 1; Length 1504;  
 Best Local Similarity 17.88; Pred. No. 32;

Matches 146; Conservative 105; Mismatches 255; Indels 316; Gaps 39;  
 QY 38 NLDRLMRELGFTLP-LDSLYSFDKSIANAVVIFGGCTGTVSOGGLIFTNHHG----- 90  
 DB 212 NVDRCY---FRSPVLNSILDIDK-----LTINDDLQLLLDRFCDFKCNV 252  
 QY 91 -----GYGAIQ-----SOSTVDHDLRD-GFVS-----RTM 115  
 DB 253 LSRDRFPRVGNGLLEIDILPQIKNREKLQHRDIHDFLEKLGDISIDPVKPYVSSARDM 312  
 QY 116 GEELPI--PGLSVKYLKIVKVTQKVEGQLKGTIDEMERL-RKAQEVQCEOLAKKENADEN 172  
 DB 313 INELTMOREELSLSKEYKEPPETKRHVSHQWSSGEAEFYKKAQH-----KSTFDG 365  
 QY 173 QLCIVEPFYSNNEVFLIVYDFKDVRFAPPSVGVGGDTDNMMPRHGTDFSVFRVY 232  
 DB 366 QIPNFENFDKNQKFSAINPYE-----ALPQ-----LWPR----- 396  
 QY 233 AGADNRPAEYSKDNKPKYVFAAVSMQGYKADDPYAMTIGFPGSTDRYLTSM---GVEDR 289  
 DB 397 -----LPQIEINNN-----SMQDKKND-----QVNASFTEYICGVNE 431  
 QY 290 IENENNRIEVRGIKQGIWEKMSADQATRIKYASKYAQSANY-----WK 334  
 DB 432 NEGKVGNIKSRSYSWLPESIASPKDSTILLDHQTQKHYHTINFSDCMTQNMASKRKL 491  
 QY 335 NSIGMNRGLARLDVIGRKRAEAFADWIRKNGKSAVVG-----DVLSSLEKA----- 382  
 DB 492 SSVSAN---KTSLLSRKR--KQVMAAGLVYKRAFYGVGPPFGYQDILNKLDEGEFPI 545  
 QY 383 -YKE-----GAKANREMYLSETL---FGTEV-----VRFAQAFAN-A 415  
 DB 546 DYKDPFTSNPVDLENKPYAVAGKRFESSTHSTRIPVQGGTGVSVYVKNKPTDFMFSSWK 605  
 QY 416 LATNPDAHAGILKSLDDKYDYLPDLDRKVLPAMLDIVRRRIPADKLPDIFKNVIDKKFK 475  
 DB 606 YALKPPTYDAV-----QKWNKVPKSMGNKKTESQISM---HTPHSKF--LYKFASDVSGK 655  
 QY 476 GDTKK-----YADFVDFKSVVPYSDKFHMLKSMDBKFA----- 510  
 DB 656 QKRKSSVHSDSLTHLTLEIHANTSDKIPDAIDEVSMIWCLEETFPPLDLDIAYEGIM 715  
 QY 511 ---KAIEK-----DPAVELSKSVIAARAIOADAM-----A 540  
 DB 716 IVHKASEDSTFPTKIQHCINEIPVMFVESEFEMFEALDVLVLLDPLDLSGFEIHNSWG 775  
 QY 541 YAIKGRKRLF-FAGLRE-----MYPG-----RALPSDAN 568  
 DB 776 YIIECRQKIHQFDIVRELARVYKCOIKYKLSDTWGYAHSSGIMITGRHMINTWRALRSVN 835  
 QY 569 FTRMSYSGTSGYEPQDGAWYNYHTTGKGVLEKQDPK-----SDEFAVOENILDLFRTK 622  
 DB 836 LTQ-----YTIESAAFNILHKLPHFESLTMNMNNAKSTTELKTVL 878  
 QY 623 NYGRYAENGOLHIAFLSNNDITGCGNSGSPVDFKNGRLIGLAF 664  
 DB 879 NY---WLSRAQINIGLLRKQDYIARNI-----EQARLIGIDF 912  
 RESULT 13  
 ACON\_RICPR  
 ID ACON\_RICPR STANDARD; PRT; 878 AA.  
 AC Q9ZCF4;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Aconitate hydratase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase).  
 GN ACONA OR RP799.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]



```

Db 204 A---ELIRPSLTIV---TTIQEATPAEGMHGAVCMQPTIVEPEGLAPEPLIARKRS 257
QY 158 EVCOELAKKENA-----DEN--OLCIV-----BPFYSNNE 185
Db 258 TAVOEVRKKSALAEVVKSKTLCGCHNEAIEFMDEETISLCILALETFLHREPSMAAPI 317
QY 186 YFLIVYDFVDRVRFAPPSSVSGKFGGDTDNMMWPHRTGDFSV----- 228
Db 318 LFKILYTV---TRLIDP-----MYPWHSTEMEVPANSVAKQWLRLVSLH 360
QY 229 -----FRVYAGADRPAEY-----SKDNKPKYKPVYFAAVSMQGYKADDYAMTIGF 273
Db 361 HLSTSAICLOLEFDTKIPRDPAFWSVVALSLADFPPELSPVYFIQILMEDLEE-----SW 413
QY 274 PGSDP---RYLTSGVGDRETNENENPRIEYVGTGKGTWK---BAMSADQA---TRIKYAS 324
Db 414 PGSVKLTKMKNLAFYIVEIPTDMYNPNPKDLVGHLETFKRYHSAISADNGITPTRAEL-- 471
QY 325 KYAOSANYKNSTGMNRLGLARLDVIGRKRAE---ERAFADWIRKNGKSAVYGDVLSLEK 381
Db 472 -----ENVITVTHVFKVQTFSSKSPVTLVEAFARWLSLHSA---DV--SLES 517
QY 382 -----AYKEGAKANREMYLSETLFCGTEVVRFAQANAL-ATNPDAHAGILKSLDDKY 434
Db 518 LLGVCTACNRALIRERDKQCITRALV--TELMQAIKFKVLHESNYVTIANMI--LQDAG 573
QY 435 KDY--LPSLDRLKVLPAMLDIVRRIPADKLPDIFKNWIDKFKGDTKRYADVFVKSVVPY 493
Db 574 EDIEVPLDDOEFNTASEAIR-----PFLF-----EVLDFIADLHWI-- 610
QY 494 SDFKPHMLKMSD-----KEKFAKAIEKOPAVELSK-----S 524
Db 611 -----AKLKRESNDALGGDLKVKLAELAI-----AVENSRSNARDCRTVIREPWLMSPPS 661
QY 525 VIAARAIQADAMAN 539
Db 662 VTQAAPSAFADSVTN 676

RESULT 15
MOKI_SCHPO STANDARD; PRT: 2410 AA.
ID Q9USK8; Q9URT5; Q9URK2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall alpha-1,3-glucan synthase moki (EC 2.4.1.183).
GN MOKI OR AGS1 OR SPC339.01C OR SPC317A7.01.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID:4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Katayama S., Dai H., Arellano M., Perez P., Toda T.;
RT "Fission yeast alpha-glucan synthase Moki localizes closely with actin
RT and play a role essential for cell morphogenesis and protein kinase C
RT function";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

```

```

RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkkaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -I- CATALYTIC ACTIVITY: UDP-glucose + (alpha-D-glucosyl-(1,3))(N) -
CC UDP + (alpha-D-glucosyl-(1,3))(N+1) -
CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB019183; BAA34054.1; -
DR EMBL; AL023781; CAAL19332.1; -
DR EMBL; AL049472; CAB39330.1; -
DR InterPro; IPR000461; Alpha_amylase.
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF00534; Glycos_transf_1; 1.
DR Cell wall; transferase; Glycosyltransferase.
FT CONFLICT 256 256 F->Y (IN REF. 2).
SQ SEQUENCE 2410 AA; 272120 MW; 8BD944BAE3A9A5C5 CRC64;

Query Match 3.0%; Score 111.5; DB 1; Length 2410;
Best Local Similarity 17.5%; Pred. No. 65;
Matches 131; Conservative 132; Mismatches 232; Indels 253; Gaps 37;

QY 88 HHCYGAIQSQSTVDHDLRDGFVSRVTGCEELPIPGLSVKY-----LRKIVK 134
Db 1355 HOKGFGAV-----GVSNKYGKRSWARYPIFWGLKKIKG 1387

QY 135 V-----TDK-VEGOLKGIITD---EMERLR-KAOEVCQELAKKENADENOLCIVEPYS 182
Db 1388 LPNPDPTDITDEIVDDKAVAITDIDPMKSKVHKRKAQEWAGLEVNEKYDYLVPVGRWS 1447

QY 183 NNEVFLIVYDFVKD-----VRMVFPAP--SSVGKFGGDTDNMMWPHR--TGDFSVFR-- 230
Db 1448 SQKGIDLIADIAPSLLESYKVLQICVGPILIDYCKFAEKDLVQKKYPTRFVSQPKFTQ 1507

QY 231 ----VYAGADNRPAEYKDNKPKYKPVYFAAVSMQ--GYADDYAMTIGPGSDTRLTLSWG 285
Db 1508 LPPYIFSGADPALI-----PSRDEPFLVAVFVRGK--ALGIGARVGGGLGQMPGMW 1557

QY 286 VEDRIENENNPRIEVRGIKQGIWKEMASDAOATRIKYASKVAQ----- 328
Db 1558 YS--VESSATPHL-LKQFEQAC-QOALSSORTLARLRSAKQRFPPVSNQAKLEALTD 1613

QY 329 -----SANYWNKNSIG-----MNRGLARLDVI-----GRKRAEERAFADWIRKNG 367
Db 1614 GCIKCSQKYGNSRSRSSFYSLIHESFSRSSEVLPTSSDITNLDKRAEE---AEMIMET 1670

QY 368 KSAVYGDVLSLEKAYKEGAK-----ANREMTYLSSETLFGCTGVVRF 410
Db 1671 PPTAEANTGAKLDRSLGSRGPGHTTDDASDGLDTIQEESMTAGDSTSGGSDISRY- 1729

```



QY 411 QFANALATNPDAIAGILKSLDDKYKOYLPSLDRKVLPAMLDIVRRRIPADKLPDIFKNVI 470  
Db 1730 ---RAERLNPDSHSPSEYFDSG-----1749  
QY 471 DKFKFGDTKKYADFVF-DKSVVPYSDKFHAMLKSMCKE-----KFAKAIEKDPAVELSKS 524  
Db 1750 DYEFDPORSYYYDDLFDDTTIRNAPSPFOMGSFDEAHAVGATFSODDLSDP-----1802  
QY 525 VIAARAIOADAM-----ANAYAIKKGKRLFFAGLREMYPGRALPSDANFTMRMS 574  
Db 1803 ----ARSVDSDSVSPPLPPFVAGSNPNARNNNPFFYGNLH-----TESSLSLA-- 1847  
QY 575 YGSIKGYEPODGAWYNYHTT-----GKGV-----LEKQDPKS--DEFAYOENILDLFRT 621  
Db 1848 -SMSGKEKRDFSLTRVEETFTDEDGQALRSFSEKLOKLNKNSKDDLCIEQYLMKRSERS 1906  
QY 622 KNYGRYA-----ENGOLHIAFLSNNDIT---GGNSGSPVEDKNGRLIGLAFDGNWEAM 671  
Db 1907 FFHERRAIKLGLOKPNKLVNELSSHSCTESESLSNGQTSYDD---IIAMTDESNTQL 1963  
QY 672 SGDIEFEPDLQRTISVDIRYVLFMIDKW 699  
Db 1964 -GDDDF-----KTIHGLKKEMLFKIYDW 1985

Search completed: May 25, 2003, 15:32:26  
Job time : 23.1927 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 10:27:28 ; Search time 149 Seconds  
(without alignments)  
151.141 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TGGNSGSPVF 10

Scoring table:  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlh  
-DB=/cgn2.1/USPTO.spool/US1008355/runat\_16052003.110459.3075/app\_query.fasta\_1.199  
-DB=N\_Geneseq\_101002 -OFT=fastp -SUFFIX=p2n.rng -MINMATCH=0 -L=LOOPEL-0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=1000 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=50  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US1008355 -CGN\_1.1.200 -runat\_16052003.110459.3075 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELEX=6 -DELEX=7

Database : N\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	54	100.0	1019	18	AAV75063	Staphylococcus aur
2	54	100.0	1558	13	AAQ27988	Protease from S. A
3	54	100.0	1586	13	AAQ27987	Protease from S. A
4	54	100.0	2139	24	AAL43635	Porphyromonas ging
5	49	90.7	849	22	AAH52485	S. epidermidis ope
6	49	90.7	936	22	ABN91429	Staphylococcus epi
7	49	90.7	3189	22	AAH54330	S. epidermidis gen
8	44	81.5	402	22	AAL17622	Human breast cance
9	44	81.5	423	22	AAL00017	Human reproductive
10	44	81.5	758	20	AAV20213	Enterococcus faeca
11	44	81.5	758	24	ABN98198	E faecalis EF10 g
12	44	81.5	888	20	AAV20212	Enterococcus faeca
13	44	81.5	888	24	ABN98197	Enterococcus faeca
14	44	81.5	2421	20	AAV80633	E faecalis EF10 g
15	44	81.5	4951	20	AAV13198	Kidney injury asso
16	44	81.5	17391	23	ABL30036	Drosophila melanog
17	43	79.6	534	24	ABK73036	Bacillus lichenifo
18	43	79.6	948	22	AAV02999	Bacillus lichenifo
19	43	79.6	1448	13	AAQ24382	Protease Blase. B
20	42	77.8	152	22	AAV79971	Nucleotide sequenc
21	42	77.8	473	23	ABV56742	Human prostate exp
22	42	77.8	1546	24	ABQ54166	Human ovarian anti
23	42	77.8	319608	21	AAH51601	Human chromosome 1
24	42	77.8	319608	22	AAV03301	Human schizophre
25	41	75.9	2130	24	AAQ87069	M. capsulatus gene
26	41	75.9	8922	22	AAK87069	Human immune/haema
27	40	74.1	370	24	ABN25944	Human OREX polynuc
28	40	74.1	450	24	ABL78727	Human ovarian canc
29	40	74.1	942	22	AAV03001	Bacillus lichenifo
30	40	74.1	954	22	AAV03003	Bacillus lichenifo
31	40	74.1	1030	22	AAH99612	Human protein enco
32	40	74.1	1127	21	AAV07536	Fusarium venenatum
33	40	74.1	1873	24	AAV19209	DNA encoding human
34	40	74.1	7441	22	AAV29979	Human lung antigen
35	40	74.1	44861	24	AAV20000	DNA encoding pyrid
36	39	72.2	240	18	AAV89059	Human prostate can
37	39	72.2	584	22	ABA31488	Probe #9954 for ge
38	39	72.2	584	22	AAK12811	Human brain expres
39	39	72.2	741	21	AAV14488	Aspergillus oryzae
40	39	72.2	1173	24	ABK32891	DNA encoding C. al
41	39	72.2	1261	20	AAV04382	Human secreted pro
42	39	72.2	1303	22	AAV26291	Pseudomonas sp typ
43	39	72.2	1775	24	ABV03070	Human breast speci
44	39	72.2	2012	21	AAV59942	Human secreted pro
45	39	72.2	2012	22	AAH33087	Human colon cancer
46	39	72.2	2063	24	ABA96872	Human flavoprotein
47	39	72.2	2107	24	ABL90832	Human polynucleoti
48	39	72.2	2109	22	AAV27076	DNA encoding nove
49	39	72.2	2136	23	ABK43738	DNA encoding novel
50	39	72.2	2166	22	AAH16791	Human CDNA sequenc
51	39	72.2	2774	11	AAQ03245	pe1B Pectin lyase
52	39	72.2	3255	22	AAI29133	Human PAC full len
53	39	72.2	3456	21	AAV37083	CDNA of a gene red
54	39	72.2	3462	21	AAV77547	Human OREX ORF3102
55	39	72.2	3678	22	AAH98399	Human EST-derived
56	39	72.2	3820	23	ABL06881	Drosophila melanog
57	39	72.2	8578	23	ABL06880	Pig alpha-1,3 gala
58	39	72.2	8989	22	AAV05297	Inhibitory clone p
59	38	70.4	93	21	AAA56105	Human single nucle
60	38	70.4	184	20	AAH86031	CDNA #1415 encoding
61	38	70.4	218	23	AAV58739	CDNA #676 encoding
62	38	70.4	222	23	AAV58000	CDNA #1285 encoding
63	38	70.4	222	23	AAV58609	CDNA #614 encoding
64	38	70.4	243	23	AAV57938	CDNA #729 encoding
65	38	70.4	253	22	AAV58053	Human polynucleoti
66	38	70.4	428	22	AAI85512	Human secreted pro
67	38	70.4	440	22	ABA08678	

c 68	38	70.4	442	22	AA183051	Human polynucleoti	141	37	68.5	3663	22	AA14352	Human mandanese su
c 69	38	70.4	448	21	AA227896	Human secreted pro	142	37	68.5	3704	22	ABQ70790	Listeria monocytog
c 70	38	70.4	454	24	ABL93826	Arabidopsis thalia	c 143	37	68.5	3740	21	AAQ26349	Human secreted pro
c 71	38	70.4	474	24	ABL82252	Human ovarian canc	c 144	37	68.5	3787	24	ABO54746	Human ovarian anti
c 72	38	70.4	531	22	AA547068	Human breast cance	145	37	68.5	4457	21	AA575360	Human ORFX ORF915
c 73	38	70.4	531	22	AA17038	Human breast cance	146	37	68.5	4503	23	AA573231	DNA encoding novel
c 74	38	70.4	531	24	ABK95103	Human breast tumor	147	37	68.5	5845	22	AA52033	Human polynucleoti
c 75	38	70.4	663	11	AAQ05615	Human breast tumor	148	37	68.5	5995	22	AA52032	Human polynucleoti
c 76	38	70.4	668	22	AA125246	Insert from PARC 1	149	37	68.5	6273	22	AA53016	Human polynucleoti
c 77	38	70.4	828	9	AA180912	Sequence encoding	c 150	37	68.5	6273	22	AA53017	Human polynucleoti
c 78	38	70.4	828	21	AA071168	Pig lung protease	c 151	37	68.5	6306	22	AAK78589	Human immune/haema
c 79	38	70.4	994	22	AA559888	Human novel cytol	152	37	68.5	6476	23	ABL03322	Drosophila melanog
c 80	38	70.4	1005	23	ABL22883	Drosophila melanog	153	37	68.5	6513	24	ABQ71039	Listeria monocytog
c 81	38	70.4	1114	22	AA559845	Human novel cytol	154	37	68.5	6692	23	AA573229	DNA encoding novel
c 82	38	70.4	1119	22	AAH68579	Human protein HP10	c 155	37	68.5	7280	24	AA573229	Human PKIN-9 cDNA
c 83	38	70.4	1719	22	AAK70665	Human immune/haema	c 156	37	68.5	7328	22	AA573229	Novel protein kina
c 84	38	70.4	1723	22	AAK70664	Human immune/haema	c 157	37	68.5	7875	22	AA573229	Human immune/haema
c 85	38	70.4	2290	22	AA526153	Human cDNA encodin	c 158	37	68.5	14186	23	ABL50559	Micromonospora car
c 86	38	70.4	2323	22	AAH18367	Human cDNA sequenc	159	37	68.5	15677	23	ABL05318	Drosophila melanog
c 87	38	70.4	2685	22	AAH14840	Human cDNA sequenc	160	37	68.5	15705	23	ABL05318	Drosophila melanog
c 88	38	70.4	2690	23	ABL22872	Drosophila melanog	c 161	37	68.5	16636	20	AA520546	Polynucleotide seq
c 89	38	70.4	2716	23	AA580398	DNA encoding novel	c 162	37	68.5	19167	22	ABA19101	Human nervous syst
c 90	38	70.4	2796	22	AA161053	Human polynucleoti	163	37	68.5	19167	22	AA105340	Human reproductive
c 91	38	70.4	3026	23	AA592486	DNA encoding novel	164	37	68.5	19167	23	ABL98209	Human testicular a
c 92	38	70.4	3061	23	AA22882	Drosophila melanog	c 165	37	68.5	46870	22	AA520546	DNA encoding human
c 93	38	70.4	3283	21	AA236154	An alternative for	c 166	37	68.5	51552	24	AA596607	C glutamic codin
c 94	38	70.4	3317	19	AAV62668	Mouse G protein co	c 167	37	68.5	349980	22	AAH68529	Bacillus lichenifo
c 95	38	70.4	3407	22	AA159267	Human polynucleoti	c 168	36	66.7	36	22	AAQ03008	Aspergillus aculea
c 96	38	70.4	3412	21	AA236152	DNA encoding cance	169	36	66.7	184	15	AAQ03820	Human foetal liver
c 97	38	70.4	3540	22	AA126620	Human breast cance	c 170	36	66.7	191	22	ABA74160	Probe #17637 for g
c 98	38	70.4	3667	23	AA570982	DNA encoding novel	c 171	36	66.7	191	22	ABA39171	Human brain expres
c 99	38	70.4	3694	22	AA526150	Human cDNA encodin	c 172	36	66.7	191	22	AAK22615	Human bone marrow
c 100	38	70.4	4427	22	AA526150	Human vesicle traf	c 173	36	66.7	191	22	AAK48783	Human bone marrow
c 101	38	70.4	4615	23	AA565973	DNA encoding novel	c 174	36	66.7	191	22	AAI26337	Probe #16270 for g
c 102	38	70.4	4617	23	AA565486	DNA encoding novel	c 175	36	66.7	191	22	AAI54612	Probe #23298 used
c 103	38	70.4	5378	24	AAK92449	Human prostate spe	c 176	36	66.7	191	24	ABS22492	Human genome-deriv
c 104	38	70.4	5571	24	ABK92245	Prostate cancer-sa	c 177	36	66.7	196	23	ABV61737	Human prostate exp
c 105	37	68.5	39	16	AA708201	Amplification prim	c 178	36	66.7	202	23	ABV61471	Human prostate exp
c 106	37	68.5	42	16	AA708198	Amplification prim	c 179	36	66.7	276	24	ABN19923	Human ORFX polynuc
c 107	37	68.5	222	24	ABS23868	Human genome-deriv	c 180	36	66.7	287	24	ABN77871	Human ORF2818 cDNA
c 108	37	68.5	301	21	AA526150	Human secreted pro	c 181	36	66.7	288	24	ABN77976	Human ORF2923 cDNA
c 109	37	68.5	387	23	ABV11629	Human prostate exp	c 182	36	66.7	290	23	ABV61542	Human prostate exp
c 110	37	68.5	388	23	ABV32274	Human prostate exp	c 183	36	66.7	330	24	ABN19255	Human ORFX polynuc
c 111	37	68.5	388	23	ABV41699	Human prostate exp	c 184	36	66.7	350	22	ABN19255	Human immune/haema
c 112	37	68.5	403	20	AAH66521	Human prostate exp	c 185	36	66.7	359	22	AAI80221	Human polynucleoti
c 113	37	68.5	410	20	AAH66521	EST clone AS167	c 186	36	66.7	376	23	ABV56400	Human prostate exp
c 114	37	68.5	414	23	AAV02460	Human polynucleoti	c 187	36	66.7	384	22	AAI93370	Human polynucleoti
c 115	37	68.5	417	22	AA191075	Human prostate exp	c 188	36	66.7	385	20	AAV89369	EST clone CK213
c 116	37	68.5	427	22	AA183458	Human polynucleoti	c 189	36	66.7	388	22	AAI88048	Human polynucleoti
c 117	37	68.5	432	22	AA189198	Human polynucleoti	c 190	36	66.7	396	22	AAF65695	Novel human polynu
c 118	37	68.5	449	22	AAI19705	Human breast cance	c 191	36	66.7	397	23	AAV06963	Human prostate exp
c 119	37	68.5	497	22	AAK61632	Human immune/haema	c 192	36	66.7	410	22	AAI83023	Human polynucleoti
c 120	37	68.5	543	23	ABV21575	Human prostate exp	c 193	36	66.7	420	23	ABV12365	Human prostate exp
c 121	37	68.5	543	23	ABV27394	Human prostate exp	c 194	36	66.7	443	23	ABV37775	Human prostate exp
c 122	37	68.5	557	24	ABS11510	Human genome-deriv	c 195	36	66.7	450	23	ABV33508	Human prostate exp
c 123	37	68.5	558	22	AAH87813	Peppermint plant o	c 196	36	66.7	453	23	ABV36906	Human prostate exp
c 124	37	68.5	633	19	AAV09282	Nucleotide sequenc	c 197	36	66.7	454	18	AAV51554	Herpes virus DNA p
c 125	37	68.5	702	24	ABQ68601	Listeria monocytog	c 198	36	66.7	461	23	ABV14510	Human prostate exp
c 126	37	68.5	720	23	AA551714	Staphylococcus aur	c 199	36	66.7	508	23	ABV47361	Human foetal liver
c 127	37	68.5	765	23	AA554542	Staphylococcus aur	c 200	36	66.7	529	22	ABA61665	Probe #7787 for ge
c 128	37	68.5	827	18	AAV74912	Staphylococcus aur	c 201	36	66.7	529	22	ABA29321	Human brain expres
c 129	37	68.5	1206	22	AAH66809	C glutamic codin	c 202	36	66.7	529	22	AAK09967	Human bone marrow
c 130	37	68.5	1231	23	AAH72501	DNA encoding novel	c 203	36	66.7	529	22	AAK35861	Probe #10763 for ge
c 131	37	68.5	1242	24	ABQ68390	Listeria monocytog	c 204	36	66.7	529	22	AAI11760	Probe #10763 for ge
c 132	37	68.5	1507	23	ABL11993	Drosophila melanog	c 205	36	66.7	529	22	AAI14577	Human genome-deriv
c 133	37	68.5	1624	24	ABQ71205	Listeria monocytog	c 206	36	66.7	529	24	ABS10091	Human prostate exp
c 134	37	68.5	1788	23	AA575513	DNA encoding novel	c 207	36	66.7	533	23	ABV27008	Human prostate exp
c 135	37	68.5	1952	21	AA264444	Human secreted pro	c 208	36	66.7	533	23	ABV29557	Human breast cance
c 136	37	68.5	2028	22	AAK69011	Human immune/haema	c 209	36	66.7	552	22	AAI22243	Human prostate exp
c 137	37	68.5	2205	24	AA562671	cDNA sequence #458	c 210	36	66.7	553	23	ABV50245	Human prostate exp
c 138	37	68.5	2229	23	AA573230	DNA encoding novel	c 211	36	66.7	566	23	ABV51720	Human prostate exp
c 139	37	68.5	2732	23	AA586858	DNA encoding novel	c 212	36	66.7	599	23	ABV41158	Human prostate exp
c 140	37	68.5	3075	24	ABK34656	Human cDNA for nov	c 213	36	66.7	599	23	ABV44848	Human prostate exp

c 214	66.7	616	22	ABL14493	Human breast cancer	287	36	66.7	7263	23	AAS69874	DNA encoding novel
c 215	66.7	629	22	ABL01477	Human reproductive	288	36	66.7	7303	24	ABK31303	Signal transductio
c 216	66.7	629	23	ABL96922	Human testicular a	289	36	66.7	7566	14	AAQ42160	Plasmid pPS0212 co
c 217	66.7	700	22	AAD08203	Human secreted pro	c 290	36	66.7	7599	22	AAF25320	Nucleotide sequenc
c 218	66.7	723	21	AAC53841	Arabidopsis thalia	c 291	36	66.7	7606	23	ABL04782	Drosophila melanog
c 219	66.7	767	22	ABL23355	Human breast cancer	c 292	36	66.7	7639	14	AAQ42159	Plasmid pJD884 con
c 220	66.7	791	24	ABL89683	Human polynucleoti	c 293	36	66.7	12687	16	AAQ98814	EAV GL gene. Equi
c 221	66.7	867	23	ABL21271	Drosophila melanog	c 294	36	66.7	12687	18	AAT70125	Equine arteritis v
c 222	66.7	990	24	AB199702	Mouse ischaemic co	c 295	36	66.7	12827	19	AAV09036	Equine arteritis v
c 223	66.7	1037	23	ABV24477	Human prostate exp	c 296	36	66.7	14194	22	AAC66932	Plant signal trans
c 224	66.7	1041	22	AAAD11166	Nitrilase DNA #1.	c 297	36	66.7	15397	18	AAT58635	T-DNA insert of 13
c 225	66.7	1110	21	AZ333308	Neisseria meningit	c 298	36	66.7	15397	18	AAT58635	T-DNA insert of 13
c 226	66.7	1142	22	AAS27010	cDNA encoding nove	c 299	36	66.7	15528	19	AAV09039	Equine arteritis v
c 227	66.7	1152	23	AAS68337	DNA encoding novel	c 300	36	66.7	17397	22	AAS36445	Human cardiovascular
c 228	66.7	1184	22	AAH94456	Human foetal cDNA,	c 301	36	66.7	19334	22	AAS36443	Human cardiovascular
c 229	66.7	1221	22	AAD02353	Human RNA metaboli	c 302	36	66.7	19345	22	AAS36444	Human cardiovascular
c 230	66.7	1242	21	AZ293415	NGSP polypeptide c	c 303	36	66.7	21606	22	AAS36233	Human cardiovascular
c 231	66.7	1266	22	AAH98567	Human protein enco	c 304	36	66.7	23075	22	ABA19112	Human nervous syst
c 232	66.7	1289	23	ABV25037	Human prostate exp	c 305	36	66.7	24593	6	AAAS0226	Sequence of opine
c 233	66.7	1289	23	ABV25356	Human prostate exp	c 306	36	66.7	24596	6	AAAS0226	Complete nucleotid
c 234	66.7	1291	15	AAZ51539	Aspergillus aculea	c 307	36	66.7	30625	24	ABK12808	Human tumour suppr
c 235	66.7	1326	21	AAZ51539	Neisseria meningit	c 308	36	66.7	38726	23	AAS59513	Propionibacterium
c 236	66.7	1347	21	AAZ51533	Neisseria meningit	c 309	36	66.7	43226	20	AAK60263	Nucleic acid sequ
c 237	66.7	1374	20	AAZ13587	Enterococcus faeca	c 310	36	66.7	49914	21	AAH81471	N. meningitidis pa
c 238	66.7	1395	21	AAZ93414	NGSP polypeptide c	c 311	36	66.7	110608	24	ABK83572	Human cDNA differe
c 239	66.7	1395	21	AAZ51538	Neisseria meningit	c 312	36	66.7	117213	19	AAV62176	HSV-2 strain SB5 C
c 240	66.7	1500	21	AAZ54510	Neisseria gonorrhe	c 313	36	66.7	154746	24	AAD25519	Human herpesvirus
c 241	66.7	1500	21	AAZ54511	Neisseria meningit	c 314	36	66.7	154746	24	AAD25519	Human herpesvirus
c 242	66.7	1500	21	AAZ54512	Neisseria meningit	c 315	36	66.7	349980	21	AAZ21607	Neisseria meningit
c 243	66.7	1500	21	AAZ33305	Neisseria meningit	c 316	36	66.7	349980	22	AAH41223	Pyrococcus abyssi
c 244	66.7	1500	21	AAZ33306	Neisseria meningit	c 317	36	66.7	349980	22	AAH41225	Pyrococcus abyssi
c 245	66.7	1500	21	AAZ33307	Neisseria meningit	c 318	36	66.7	349980	22	AAH68527	C glutamicum codin
c 246	66.7	1595	19	AAV32012	Mus musculus cathe	c 319	36	66.7	349980	22	AAH68530	C glutamicum codin
c 247	66.7	1629	20	AAZ22018	Human channel-rela	c 320	36	66.7	513445	22	AAI61373	Soybean 318013 reg
c 248	66.7	1629	20	AAH66103	C glutamicum codin	c 321	36	66.7	580073	18	AAT58840	Mycoplasma genital
c 249	66.7	1931	21	AAH16149	Human prostate can	c 322	36	66.7	1437668	21	AAA81490	N. meningitidis B
c 250	66.7	2025	22	AAH71273	Corynebacterium g1	c 323	35	64.8	45	18	AAT74897	JEMSI used in cons
c 251	66.7	2295	23	AAH52069	DNA encoding novel	c 324	35	64.8	62	22	AAH73508	HGF nucleic acid 1
c 252	66.7	2370	22	AAH52044	Mycobacterium tube	c 325	35	64.8	62	22	AAH73509	HGF nucleic acid 1
c 253	66.7	2433	24	AAH33665	Human TRIC-20 cDN	c 326	35	64.8	100	16	AAQ84567	HSV type 1 IE4 and
c 254	66.7	2678	23	ABV23462	Human prostate exp	c 327	35	64.8	160	22	ABA88769	Escherichia coli p
c 255	66.7	2678	23	ABV29320	Human prostate exp	c 328	35	64.8	164	23	ABV60939	Human prostate exp
c 256	66.7	2764	21	AAH77932	Human cancer assoc	c 329	35	64.8	168	24	ABU83905	Human ovarian can
c 257	66.7	2905	24	ABK14683	cDNA encoding nove	c 330	35	64.8	182	23	ABV49466	Human prostate exp
c 258	66.7	2916	24	ABL41975	DNA sequence of H4	c 331	35	64.8	192	22	ABA69680	Human foetal liver
c 259	66.7	2981	23	ABL21270	Drosophila melanog	c 332	35	64.8	192	24	ABS17947	Human genome-deriv
c 260	66.7	3191	21	AAA16647	Human secreted pro	c 333	35	64.8	204	24	ABL36817	Human colon tumour
c 261	66.7	3364	22	ABA09220	Human TRAP150 homo	c 334	35	64.8	240	22	ABA50851	Human breast cell
c 262	66.7	3410	23	ABL09385	Drosophila melanog	c 335	35	64.8	240	22	ABA68824	Human foetal liver
c 263	66.7	3423	23	ABL24497	Drosophila melanog	c 336	35	64.8	240	22	ABA35777	Probe #14243 for g
c 264	66.7	3547	21	AAZ29663	Mouse Serine prote	c 337	35	64.8	240	22	AAK17160	Human brain expres
c 265	66.7	3611	22	AAH18179	Nucleotide sequenc	c 338	35	64.8	240	22	AAK42945	Human bone marrow
c 266	66.7	3618	24	ABL56484	Drosophila melanog	c 339	35	64.8	240	22	AAI23709	Probe #13642 for g
c 267	66.7	3620	23	ABL02471	Drosophila melanog	c 340	35	64.8	240	22	AAI49022	Probe #17708 used
c 268	66.7	3656	20	AAZ23299	Human SPA-1 cDNA.	c 341	35	64.8	240	22	AAI09326	Probe #9317 used t
c 269	66.7	3774	23	ABL21474	Drosophila melanog	c 342	35	64.8	240	22	ABS17007	Human genome-deriv
c 270	66.7	3958	13	AAQ29230	The abf-A gene of	c 343	35	64.8	245	21	AAK28314	Human secreted pro
c 271	66.7	4260	23	ABL04783	Drosophila melanog	c 344	35	64.8	247	22	AAK23321	Human brain expres
c 272	66.7	4488	23	ABL02548	Drosophila melanog	c 345	35	64.8	279	24	ABL75715	Corn tassal-deriv
c 273	66.7	4522	21	AAH76440	Human ORFX ORF195	c 346	35	64.8	293	24	ABN20799	Human ORFX polynuc
c 274	66.7	4740	23	ABV21724	Human prostate exp	c 347	35	64.8	304	23	ABV52744	Human prostate exp
c 275	66.7	4740	23	ABV27546	Human prostate exp	c 348	35	64.8	306	22	AAK45667	Human bone marrow
c 276	66.7	4947	21	AAZ60627	Nucleotide sequenc	c 349	35	64.8	306	22	AAI51593	Probe #20279 used
c 277	66.7	5399	13	AAQ25706	pDE108. Synthetic	c 350	35	64.8	306	24	ABN19939	Human genome-deriv
c 278	66.7	5410	23	ABL09984	Drosophila melanog	c 351	35	64.8	306	24	ABN26504	Human ORFX polynuc
c 279	66.7	5561	22	AAI59378	Human polynucleoti	c 352	35	64.8	309	24	ABQ90908	M. capsulatus gene
c 280	66.7	6077	23	ABL18984	Drosophila melanog	c 353	35	64.8	324	22	AAI90870	Human polynucleoti
c 281	66.7	6222	23	ABL24496	Drosophila melanog	c 354	35	64.8	331	23	ABV48895	Human prostate exp
c 282	66.7	6415	23	ABL02470	Drosophila melanog	c 355	35	64.8	349	22	AAI37664	Human musculoskele
c 283	66.7	6555	15	AAQ53874	Plasmid pVE144 use	c 356	35	64.8	350	23	ABV61145	Human prostate exp
c 284	66.7	6854	19	AAT76903	S. glaucescens pst	c 357	35	64.8	352	23	ABV07453	Human prostate exp
c 285	66.7	7050	14	AAQ40419	Sequence of pTE4.	c 358	35	64.8	353	22	AAH71763	Human cervical can
c 286	66.7	7097	23	ABL21272	Drosophila melanog	c 359	35	64.8	358	22	AAI80857	Human polynucleoti

c 360	35	64.8	363	22	AAF65766	Novel human polynu	c 433	35	64.8	766	23	AAS94191	DNA encoding novel
c 361	35	64.8	365	23	ABN12077	Human prostate exp	434	35	64.8	778	21	AAF298149	Human signal pepti
c 362	35	64.8	367	24	ABN97143	Gene #3641 used to	435	35	64.8	783	21	AAF12814	Aspergillus oryzae
c 363	35	64.8	375	23	ABV17542	Human prostate exp	436	35	64.8	795	23	ABV30278	Human prostate exp
c 364	35	64.8	378	16	AO800358	HCR-15 contg. a DN	437	35	64.8	816	21	AAZ53090	Neisseria meningit
c 365	35	64.8	381	22	AAI182385	Human polynucleoti	438	35	64.8	830	20	AAZ33519	Human H19 promoter
c 366	35	64.8	387	23	ABV61863	Human prostate exp	439	35	64.8	830	22	AAZ72969	Human H19 promoter
c 367	35	64.8	389	22	AAZ44889	Human contig polyn	440	35	64.8	830	22	AAZ21350	Human H19 promoter
c 368	35	64.8	391	22	ABV55819	Human prostate exp	441	35	64.8	837	21	AAZ09000	Fusarium venenatum
c 369	35	64.8	401	22	AAI182049	Human polynucleoti	442	35	64.8	839	22	AAH70564	Human cervical can
c 370	35	64.8	405	23	ABV49051	Human prostate exp	443	35	64.8	841	20	AAZ76502	Human WISP-2 prote
c 371	35	64.8	407	22	AAI186992	Human polynucleoti	444	35	64.8	850	18	AAZ62839	C. hirsutus GPD ge
c 372	35	64.8	407	23	ABV33222	Human prostate exp	445	35	64.8	871	18	AAZ60698	Viral infection ge
c 373	35	64.8	412	14	AAO60971	Human brain Expres	446	35	64.8	888	22	AAH72572	Human cervical can
c 374	35	64.8	412	23	ABV02908	Human prostate exp	447	35	64.8	917	20	AAZ27343	Human secreted pro
c 375	35	64.8	414	22	AAI183638	Human polynucleoti	448	35	64.8	940	11	AAQ06394	Exon 2 of porcine
c 376	35	64.8	414	22	AAI183988	Human polynucleoti	449	35	64.8	990	19	AAV49593	Human osterocarcin
c 377	35	64.8	416	22	AAK59172	Human immune/haema	450	35	64.8	993	23	AAZ90290	DNA encoding novel
c 378	35	64.8	417	24	ABL78153	Human ovarian canc	451	35	64.8	1132	19	AAV56045	N. excelsior water
c 379	35	64.8	421	22	AAI35217	Human musculoskele	452	35	64.8	1140	24	ABA95640	Acid phosphatase 1
c 380	35	64.8	423	23	ABV37387	Human prostate exp	453	35	64.8	1146	21	AAC42449	Arabidopsis thalia
c 381	35	64.8	424	23	ABL17993	Drosophila melanog	454	35	64.8	1164	23	AAZ54059	Pseudomonas aerugi
c 382	35	64.8	425	24	ABL51098	Human DL exon 1 SE	455	35	64.8	1215	22	AAZ86987	Nucleotide sequenc
c 383	35	64.8	429	22	AAI93485	Human polynucleoti	456	35	64.8	1215	24	ABU95649	Human angiogeneses
c 384	35	64.8	445	23	ABV61343	Human prostate exp	457	35	64.8	1215	24	ABL88160	Human PRO1361 cDNA
c 385	35	64.8	449	22	ABA45731	Human breast cell	458	35	64.8	1224	19	AAV34165	Human secreted pro
c 386	35	64.8	449	22	ABA56240	Human foetal liver	459	35	64.8	1227	20	AAV63788	Recombinant gene e
c 387	35	64.8	449	22	ABA25980	Probe #4346 for ge	460	35	64.8	1238	18	AAV74543	Staphylococcus aur
c 388	35	64.8	449	22	AAK04423	Human brain expres	461	35	64.8	1257	20	AAZ28435	EGF-like homologue
c 389	35	64.8	449	22	AAK29919	Human bone marrow	462	35	64.8	1286	21	AAA30048	Human PRO261 nucle
c 390	35	64.8	449	22	AAI14509	Probe #4442 for ge	463	35	64.8	1286	22	AAZ21403	Human <DNA sequenc
c 391	35	64.8	449	22	AAI35984	Probe #4570 used t	464	35	64.8	1266	22	AAZ60368	PRO261 coding sequ
c 392	35	64.8	449	22	AAI04334	Probe #4325 used t	465	35	64.8	1266	22	AAZ97451	Human angiogeneses
c 393	35	64.8	449	24	ABS04487	Human genome-deriv	466	35	64.8	1267	21	AAZ77566	Human PRO261 cDNA
c 394	35	64.8	452	22	AAK32643	Human bone marrow	467	35	64.8	1273	23	ABV24884	Human prostate exp
c 395	35	64.8	452	22	AAI38458	Probe #7144 used t	468	35	64.8	1280	21	ABO62495	Mycobacterium bowi
c 396	35	64.8	452	22	ABS07434	Human genome-deriv	469	35	64.8	1285	19	AAZ29260	Human connective t
c 397	35	64.8	453	23	ABV10991	Human prostate exp	470	35	64.8	1288	23	ABL05541	Drosophila melanog
c 398	35	64.8	463	22	AAK63571	Human immune/haema	471	35	64.8	1293	20	AAZ76486	Human WISP-2 prote
c 399	35	64.8	482	22	ABA57123	Human foetal liver	472	35	64.8	1309	22	AAZ82114	Connective tissue
c 400	35	64.8	482	24	ABS05452	Human genome-deriv	473	35	64.8	1316	19	AAV49592	Human Osterocarcin
c 401	35	64.8	489	22	AAI186321	Human polynucleoti	474	35	64.8	1337	22	AAH46952	Human secreted pro
c 402	35	64.8	490	23	ABV47336	Human prostate exp	475	35	64.8	1341	24	ABQ15050	Oligonucleotide fo
c 403	35	64.8	491	21	AAO00679	Human Hydrolase pr	476	35	64.8	1341	24	ABQ15051	Oligonucleotide fo
c 404	35	64.8	510	22	AAI171921	Human growth facto	477	35	64.8	1351	23	ABL05931	Drosophila melanog
c 405	35	64.8	518	23	ABV61888	Human prostate exp	478	35	64.8	1352	22	AAH46936	Human secreted pro
c 406	35	64.8	520	23	ABV06568	Human prostate exp	479	35	64.8	1411	14	AAO41260	Encodes repressor
c 407	35	64.8	529	22	AAK10628	Human brain expres	480	35	64.8	1411	15	AAQ72217	Dialkylglycine dec
c 408	35	64.8	538	21	AAZ82263	N. meningitidis pa	481	35	64.8	1444	24	ABL61072	Histidine tRNA syn
c 409	35	64.8	560	22	AAH08778	Human cDNA clone (	482	35	64.8	1483	22	AAZ93752	Human cDNA encodin
c 410	35	64.8	562	21	AAZ14808	Aspergillus oryzae	483	35	64.8	1498	22	AAZ76022	Human ORFX ORF1577
c 411	35	64.8	574	22	AAZ93944	Primer specific fo	484	35	64.8	1522	20	AAZ16595	Human growth facto
c 412	35	64.8	615	24	ABO44722	Oligonucleotide fo	485	35	64.8	1622	21	AAZ98147	Human signal pepti
c 413	35	64.8	615	23	ABO44723	Oligonucleotide fo	486	35	64.8	1630	21	AAZ49249	Serine-threonine k
c 414	35	64.8	622	23	AAZ68338	DNA encoding novel	487	35	64.8	1639	21	AAA30239	Mammalian apoptosi
c 415	35	64.8	634	22	AAH07947	Human cDNA clone (	488	35	64.8	1639	22	AAZ09571	Human transporter
c 416	35	64.8	651	22	AAH05665	Human cDNA clone (	489	35	64.8	1650	20	AAZ32320	A. thaliana EL4 DN
c 417	35	64.8	665	24	ABO60605	Human colon cancer	490	35	64.8	1661	22	AAH17023	Human cDNA sequenc
c 418	35	64.8	673	22	AAI35634	Human musculoskele	491	35	64.8	1670	22	ABA48532	Human breast cell
c 419	35	64.8	678	21	AAI12747	Aspergillus oryzae	492	35	64.8	1670	22	ABA66438	Human foetal liver
c 420	35	64.8	682	24	ABQ38744	Oligonucleotide fo	493	35	64.8	1670	22	ABA33498	Probe #11964 for g
c 421	35	64.8	684	24	ABQ38745	Oligonucleotide fo	494	35	64.8	1670	22	AAK14856	Human brain expres
c 422	35	64.8	684	24	ABK39403	DNA encoding lung	495	35	64.8	1670	22	AAK40591	Human bone marrow
c 423	35	64.8	690	24	ABL42311	Human colon tumour	496	35	64.8	1670	22	AAI21359	Probe #11292 for g
c 424	35	64.8	699	21	AAZ80206	Human colon cancer	497	35	64.8	1670	22	AAI46639	Probe #15325 used
c 425	35	64.8	700	18	AAV06133	Viral infection ge	498	35	64.8	1670	22	AAI07053	Probe #7044 used t
c 426	35	64.8	700	22	AAZ92920	Human inflammatory	499	35	64.8	1670	22	ABSI4566	Human genome-deriv
c 427	35	64.8	725	21	AAZ97446	Human prostate can	500	35	64.8	1686	23	ABS14566	DNA encoding novel
c 428	35	64.8	730	22	AAZ68225	Human lung tumour	501	35	64.8	1728	24	ABK65233	Arabidopsis cDNA e
c 429	35	64.8	730	24	ABK38136	cDNA encoding clon	502	35	64.8	1752	23	AAZ72292	DNA encoding novel
c 430	35	64.8	738	20	AAZ76501	Human WISP-2 prote	503	35	64.8	1791	17	AAZ04403	Heat shock protein
c 431	35	64.8	748	20	AAZ98857	Human validated ca	504	35	64.8	1810	16	AAZ90210	Brassica condensin
c 432	35	64.8	750	20	AAZ76487	Human WISP-2 prote	505	35	64.8	1827	21	AAZ50067	Wheat adenosine-5'

c 506	35	64.8	1859	22	AAD10126	Mouse IGFBP-5 cDNA	579	35	64.8	4219	23	ABV22856	Human prostate exp
c 507	35	64.8	1859	23	ABL35059	Murine cDNA isolat	580	35	64.8	4219	23	ABV24793	Human prostate exp
c 508	35	64.8	1893	24	AAS84509	DNA encoding novel	581	35	64.8	4219	23	ABV25865	Human prostate exp
c 509	35	64.8	1893	23	AAS85072	DNA encoding novel	582	35	64.8	4219	23	ABV27704	Human prostate exp
c 510	35	64.8	1893	23	AAS86642	DNA encoding novel	583	35	64.8	4219	23	ABV28685	Human prostate exp
c 511	35	64.8	1896	21	AZA24827	S. coelicolor YesW	c 584	35	64.8	4235	24	ABL51023	Human DL encoding
c 512	35	64.8	1949	21	AAC33574	Arabidopsis thalia	585	35	64.8	4296	21	AA254326	Neisseria meningit
c 513	35	64.8	1957	24	ABA91317	Human colon specif	586	35	64.8	4319	17	AAT17215	Adhesion and penet
c 514	35	64.8	1959	22	AAS26884	Human cDNA encodin	587	35	64.8	4350	20	AAZ12252	Neisseria meningit
c 515	35	64.8	1965	22	ABA43417	Human breast cell	588	35	64.8	4365	21	AAZ238917	Neisseria meningit
c 516	35	64.8	1965	22	ABA53861	Human foetal liver	c 589	35	64.8	4370	24	ABK84613	Human cDNA differe
c 517	35	64.8	1965	22	ABA23608	Probe #2074 for ge	590	35	64.8	4374	20	AAZ12251	Neisseria meningit
c 518	35	64.8	1965	22	AAK02125	Human brain expres	591	35	64.8	4374	21	AAF21587	N. meningitidis pa
c 519	35	64.8	1965	22	AAK27574	Human bone marrow	592	35	64.8	4374	21	AAZ1302	N. meningitidis pa
c 520	35	64.8	1965	22	AAI12159	Probe #2092 for ge	593	35	64.8	4374	21	AAZ238918	Neisseria meningit
c 521	35	64.8	1965	22	AAI33504	Probe #2190 used t	c 594	35	64.8	4381	24	ABK82001	cDNA encoding huma
c 522	35	64.8	1965	22	AAI02071	Probe #2062 used t	595	35	64.8	4407	20	AAZ12253	Neisseria gonorrh
c 523	35	64.8	1965	24	ABS02050	Human genome-deriv	596	35	64.8	4457	23	ABL06791	Drosophila melanog
c 524	35	64.8	1975	23	AAS87646	DNA encoding novel	c 597	35	64.8	4627	24	ABK27797	Human colon cancer
c 525	35	64.8	1995	22	AAH94367	Human foetal cDNA,	598	35	64.8	4756	23	ABL10784	Drosophila melanog
c 526	35	64.8	2008	21	AAC49056	Arabidopsis thalia	599	35	64.8	4784	22	AAH98417	Murine EST-derived
c 527	35	64.8	2015	21	AAA58078	cDNA encoding muri	c 600	35	64.8	4791	22	AAS22677	Human cDNA encodin
c 528	35	64.8	2033	24	ABK13620	S. aureus gene for	601	35	64.8	4794	23	ABL10790	Drosophila melanog
c 529	35	64.8	2088	23	AAS72120	DNA encoding novel	602	35	64.8	5015	22	AAAL05167	Human reproductive
c 530	35	64.8	2099	24	ABQ69021	Listeria monocytog	603	35	64.8	5015	23	ABL98054	Human testicular a
c 531	35	64.8	2125	23	ABL11241	Drosophila melanog	c 604	35	64.8	5076	24	ABL61769	Colon adenocarcino
c 532	35	64.8	2144	23	AAS72121	DNA encoding novel	605	35	64.8	5091	11	AAQ06164	The iqa gene, enco
c 533	35	64.8	2172	22	AAH15066	Human cDNA sequenc	606	35	64.8	5157	23	ABL25867	Drosophila melanog
c 534	35	64.8	2181	22	ABA82978	Human transcriptio	c 607	35	64.8	5225	23	ABL27574	Drosophila melanog
c 535	35	64.8	2208	24	AAAL0121	Isoprenoid related	608	35	64.8	5234	23	ABL11240	Drosophila melanog
c 536	35	64.8	2219	22	AAH99736	Human protein enco	c 609	35	64.8	5344	23	AAS84377	DNA encoding novel
c 537	35	64.8	2223	22	AAS33346	DNA encoding human	c 610	35	64.8	5346	21	ABL09341	Drosophila melanog
c 538	35	64.8	2359	22	AAS28798	Human immunoglob	c 611	35	64.8	5481	22	AAF31638	Mycobacterium tube
c 539	35	64.8	2370	22	ABA06478	Human cDNA SEQ ID	c 612	35	64.8	5485	22	AAH52061	Mycobacterium tube
c 540	35	64.8	2379	24	ABQ72563	Human MBDT encodin	c 613	35	64.8	5583	22	ABA08698	Human breast cance
c 541	35	64.8	2570	23	AAS84346	DNA encoding novel	c 614	35	64.8	5583	19	AAV57163	Partial human Notc
c 542	35	64.8	2627	22	AAH17965	Human cDNA sequenc	c 615	35	64.8	5646	23	ABL02910	Drosophila melanog
c 543	35	64.8	2739	20	AAZ11646	Tomato leaf curi v	616	35	64.8	5735	23	ABL23613	Drosophila melanog
c 544	35	64.8	2760	24	ABK32901	DNA encoding human	617	35	64.8	5735	21	AAC55314	Human activation-i
c 545	35	64.8	2782	24	ABA97767	Arabidopsis thalia	c 618	35	64.8	6619	20	AAV63789	Plasmid pKW2768 e
c 546	35	64.8	2784	23	ABL17992	Drosophila melanog	c 619	35	64.8	6647	22	ABA20362	Human nervous syst
c 547	35	64.8	2821	22	AAH89971	Human bone marrow	c 620	35	64.8	6647	22	AAAL05754	Human reproductive
c 548	35	64.8	2842	21	AAD00334	Rice raffinose syn	621	35	64.8	6696	22	AAS36321	Human cardiovascular
c 549	35	64.8	2961	22	AAS63190	Human purified sec	622	35	64.8	6696	22	AAAL05168	Human reproductive
c 550	35	64.8	2970	21	AAZ87102	Human thyroid rece	623	35	64.8	6696	23	ABL98055	Human testicular a
c 551	35	64.8	3081	23	ABL27575	Drosophila melanog	624	35	64.8	6829	23	ABL14618	Drosophila melanog
c 552	35	64.8	3314	22	AAS26311	Human cDNA encodin	c 625	35	64.8	7125	23	ABL26652	Drosophila melanog
c 553	35	64.8	3320	22	AAS25852	Human cDNA encodin	c 626	35	64.8	7619	22	AAS44934	DNA encoding nove
c 554	35	64.8	3365	23	AAS92466	DNA encoding novel	c 627	35	64.8	7654	22	ABA88991	Escherichia coli p
c 555	35	64.8	3464	22	AAH15732	Human cDNA sequenc	c 628	35	64.8	7800	21	AAZ53860	Neisseria meningit
c 556	35	64.8	3489	24	ABL62107	Colon adenocarcino	c 629	35	64.8	7868	24	ABQ71053	Listeria monocytog
c 557	35	64.8	3566	18	AAT69938	Coprinus cinereus	c 630	35	64.8	8091	19	AAV57001	Human Notch3 cDNA
c 558	35	64.8	3580	21	AAC58112	Human PRO1295 nucl	c 631	35	64.8	8267	23	ABL09340	Drosophila melanog
c 559	35	64.8	3580	22	AAZ37037	Human PRO1295 (UNQ	c 632	35	64.8	8515	22	AAS36319	Human cardiovascular
c 560	35	64.8	3580	22	AAZ54239	DNA encoding prote	633	35	64.8	8515	22	AAAL05166	Human reproductive
c 561	35	64.8	3707	23	AAS78011	DNA encoding novel	634	35	64.8	8515	23	ABL98053	Human testicular a
c 562	35	64.8	3707	23	AAS87645	DNA encoding novel	635	35	64.8	8954	23	ABL23612	Drosophila melanog
c 563	35	64.8	3707	23	AAS88849	DNA encoding novel	c 636	35	64.8	9051	23	ABL06790	Drosophila melanog
c 564	35	64.8	3734	22	AAS33201	DNA encoding human	637	35	64.8	9404	22	AAS36326	Human cardiovascular
c 565	35	64.8	3775	22	AAI61061	Human polynucleoti	638	35	64.8	9404	22	AAAL05171	Human reproductive
c 566	35	64.8	3813	23	ABL14619	Drosophila melanog	639	35	64.8	9404	23	ABL98058	Human testicular a
c 567	35	64.8	3902	23	ABL05930	Drosophila melanog	640	35	64.8	9796	22	AAS36325	Human cardiovascular
c 568	35	64.8	3952	22	AAAL03258	Human reproductive	641	35	64.8	9796	22	AAAL05170	Human reproductive
c 569	35	64.8	3955	22	AAAL03257	Human reproductive	642	35	64.8	9796	22	ABL98057	Human testicular a
c 570	35	64.8	3969	23	ABL17984	Drosophila melanog	643	35	64.8	9804	22	AAS36329	Human cardiovascular
c 571	35	64.8	4034	23	AAS36328	Human cardiovascular	644	35	64.8	9804	22	AAAL05173	Human reproductive
c 572	35	64.8	4034	23	ABL05540	Drosophila melanog	645	35	64.8	9804	23	ABL98060	Human testicular a
c 573	35	64.8	4113	22	ABA89172	Escherichia coli p	646	35	64.8	9805	22	AAS36324	Human cardiovascular
c 574	35	64.8	4128	22	ABA88994	Escherichia coli p	647	35	64.8	9805	22	AAAL05169	Human reproductive
c 575	35	64.8	4167	23	AAS70130	DNA encoding novel	648	35	64.8	9805	23	ABL98056	Human testicular a
c 576	35	64.8	4167	24	ABL69108	Kidney cancer rela	649	35	64.8	9831	22	AAS36327	Human cardiovascular
c 577	35	64.8	4182	22	AAAF60648	Recombinant Herpes	c 650	35	64.8	9894	23	ABL25866	Drosophila melanog
c 578	35	64.8	4219	23	ABV21874	Human prostate exp	c 651	35	64.8	10594	20	AAZ13377	Enterococcus faeca

652	35	64.8	10642	24	ABA95225	Human protease pep	725	34	63.0	70	18	AAT78726	SELEX generated l1
653	35	64.8	10664	22	AAF80046	Nucleotide sequenc	c 726	34	63.0	110.	21	AAC20688	Human secreted pro
654	35	64.8	11204	21	AAC55339	Human activation-i	c 727	34	63.0	134	21	AAC10483	Human secreted pro
655	35	64.8	11607	23	ABL13693	Drosophila melanog	c 728	34	63.0	192	13	AAV28178	Human T lymphocyte
656	35	64.8	12001	16	AAQ76213	HSV L/ST region.	c 729	34	63.0	195	20	AAV71171	PS108 gene-specific
657	35	64.8	12157	22	ABA21177	Human nervous syst	c 730	34	63.0	195	21	AAC63196	Human prostate-rel
658	35	64.8	13116	24	ABA99771	Murine capn12 DNA.	c 731	34	63.0	195	22	AAS07592	Human secreted exp
659	35	64.8	13255	22	AAK76842	Human immune/haema	c 732	34	63.0	205	21	AAK44392	Staphylococcus aur
660	35	64.8	13591	22	AAK76842	Nucleotide sequenc	c 733	34	63.0	212	18	AAV76056	Bacillus clausii g
661	35	64.8	13862	22	ABA08208	Human ovarian and	c 734	34	63.0	213	24	ABK80172	Human cardiovascular
662	35	64.8	13862	22	AAU02789	Human reproductive	c 735	34	63.0	221	22	AAS35476	Human secreted pro
663	35	64.8	13862	22	AAU07316	Human reproductive	c 736	34	63.0	221	22	AAK62822	Human immune/haema
664	35	64.8	14392	23	ABL13692	Drosophila melanog	c 737	34	63.0	222	21	AAK61129	Human prostate exp
665	35	64.8	14887	22	ABA08597	Human LDL receptor	c 738	34	63.0	228	23	ABV56911	Human immune/haema
666	35	64.8	14896	24	ABK24096	Human alpha2 macro	c 740	34	63.0	265	22	AAK69168	Human immune/haema
667	35	64.8	14898	22	ABK25224	Human polynucleoti	c 741	34	63.0	265	22	AAK69169	Bacillus clausii g
668	35	64.8	16019	22	AAS35884	Human cardiovascular	c 742	34	63.0	267	24	ABK0632	Human prostate exp
669	35	64.8	16526	21	AAK81472	N. meningitidis pa	c 743	34	63.0	274	24	ABV61286	Human prostate exp
670	35	64.8	16579	22	ABA20363	Human nervous syst	c 744	34	63.0	274	24	ABL71573	Corn tassel-derive
671	35	64.8	16579	22	AAU05755	Human reproductive	c 745	34	63.0	276	16	AAQ96370	TCR Vbeta5.3 cDNA
672	35	64.8	17335	23	ABK42393	Genomic sequence #	c 746	34	63.0	279	13	ABV61517	Human prostate,exp
673	35	64.8	17651	23	AAS59519	Propionibacterium	c 747	34	63.0	284	20	ABV89481	EST clone COL70.
674	35	64.8	19882	23	ABK42394	Genomic sequence #	c 748	34	63.0	296	21	AAA44370	Human secreted exp
675	35	64.8	22400	22	AAS35885	Human cardiovascular	c 749	34	63.0	297	21	AAC02403	Human secreted pro
676	35	64.8	26778	21	AAK81477	N. meningitidis pa	c 750	34	63.0	298	21	AAZ42600	Human 5' EST sola
677	35	64.8	28676	22	AAK80349	Human immune/haema	c 751	34	63.0	300	20	AAZ12812	Human gene express
678	35	64.8	30078	21	AAK81520	N. meningitidis pa	c 752	34	63.0	304	23	ABK60991	Human prostate exp
679	35	64.8	30078	21	AAK81520	N. meningitidis pa	c 753	34	63.0	312	24	ABK76563	Bacillus lichenifo
680	35	64.8	44377	18	AAT78508	Platenolide syntha	c 754	34	63.0	313	20	AAK91217	T. gondii immunoge
681	35	64.8	46870	22	AAK82925	S. enterica serova	c 755	34	63.0	327	23	AAS65961	DNA encoding novel
682	35	64.8	48254	22	ABA89141	Escherichia coli p	c 756	34	63.0	328	22	AAS37125	Novel human diagno
683	35	64.8	48345	22	AAK89142	Escherichia coli p	c 757	34	63.0	328	22	AAS37125	DNA encoding novel
684	35	64.8	49914	21	AAK81471	N. meningitidis pa	c 758	34	63.0	342	23	AAK71184	DNA encoding novel
685	35	64.8	54297	22	AAK70625	Human immune/haema	c 759	34	63.0	348	24	ABN75923	Human DNA/RNA bind
686	35	64.8	76798	24	ABN97454	Gene #3952 used to	c 760	34	63.0	351	22	AAI91898	Human polynucleoti
687	35	64.8	92638	24	ABN97454	Human osteoblast d	c 761	34	63.0	357	22	AAC07061	Human secreted pro
688	35	64.8	136284	24	ABK83575	Human cDNA differe	c 762	34	63.0	360	22	AAI86956	Human polynucleoti
689	35	64.8	160271	22	AAF85750	Bipolar affective	c 763	34	63.0	361	22	AAC02484	Human secreted pro
690	35	64.8	160271	22	AAF85756	Human chromosome 1	c 764	34	63.0	373	22	AAI85748	Human polynucleoti
691	35	64.8	160271	22	AAS04858	Human chromosome 1	c 765	34	63.0	374	22	AAK54147	Human polynucleoti
692	35	64.8	160271	22	AAS04858	Human chromosome 1	c 766	34	63.0	374	22	AAF93725	Murine transport a
693	35	64.8	160271	22	AAS06667	160kb fragment of	c 767	34	63.0	379	22	AAI86187	cDNA encoding SKT
694	35	64.8	160271	22	AAH40997	Human chromosome 1	c 768	34	63.0	380	22	AAI86187	Human polynucleoti
695	35	64.8	160271	22	AAH40997	Human chromosome 1	c 769	34	63.0	381	24	ABL85950	Human T lymphocyte
696	35	64.8	160271	22	AAH40997	Human chromosome 1	c 770	34	63.0	381	24	ABL85950	Human ovarian canc
697	35	64.8	172325	21	AAF85116	Nucleotide sequenc	c 771	34	63.0	382	24	ABL80915	Human ovarian canc
698	35	64.8	215980	22	AAF21613	Neisseria meningit	c 772	34	63.0	387	22	AAI91755	Human polynucleoti
699	35	64.8	240825	22	AAF24497	Complementary stra	c 773	34	63.0	391	23	ABV00477	Human prostate exp
700	35	64.8	249487	24	ABN85733	Human PG-3 gene.	c 774	34	63.0	392	22	AAI81502	Eucalyptus grandis
701	35	64.8	302250	24	ABL67703	Mouse genomic regi	c 775	34	63.0	396	21	AAI79420	Human breast cell
702	35	64.8	349980	21	AAF21607	Oesophagus cancer	c 776	34	63.0	396	22	ABA46008	Human breast cell
703	35	64.8	349980	21	AAF21608	Neisseria meningit	c 777	34	63.0	396	22	ABA31124	Human foetal liver
704	35	64.8	349980	21	AAF21612	Neisseria meningit	c 778	34	63.0	396	22	ABA36535	Human foetal liver
705	35	64.8	349980	21	AAF21612	Neisseria meningit	c 779	34	63.0	396	22	ABA69113	Probe #4628 for ge
706	35	64.8	349980	21	AAH41223	Pyrococcus abyssi	c 780	34	63.0	396	22	ABA36050	Probe #14516 for g
707	35	64.8	349980	22	AAH41226	Pyrococcus abyssi	c 781	34	63.0	396	22	ABA36050	Probe #14516 for g
708	35	64.8	534720	19	AAV30458	Rhizobium species	c 782	34	63.0	396	22	AAK04694	Human brain expres
709	35	64.8	536165	19	AAV30459	Rhizobium species	c 783	34	63.0	396	22	AAK04694	Human brain expres
710	35	64.8	580073	18	AAF58840	Mycoplasma genital	c 784	34	63.0	396	22	AAK30209	Human bone marrow
711	35	64.8	611590	21	AAF22303	Arabidopsis thalia	c 785	34	63.0	396	22	AAK30209	Human bone marrow
712	35	64.8	837096	21	AAK81489	N. meningitidis pa	c 786	34	63.0	396	22	AAI14813	Probe #4746 for ge
713	35	64.8	143768	21	AAK81490	N. meningitidis B	c 787	34	63.0	396	22	AAI24004	Probe #13937 for g
714	35	64.8	403765	22	AAI99683	Mycobacterium tube	c 788	34	63.0	396	22	AAI36171	Probe #4857 used t
715	35	64.8	441529	22	AAI99682	Mycobacterium tube	c 789	34	63.0	396	22	AAI49304	Probe #17990 used t
716	34.5	63.9	1907	22	AAF58406	Murine mOCIL2 clon	c 790	34	63.0	396	22	AAI04604	Probe #4595 used t
717	34.5	63.9	2340	24	ABA90871	Bacillus anthracis	c 791	34	63.0	396	22	AAI09592	Probe #9583 used t
718	34.5	63.9	10906	19	AAV31249	E. coli J96 pathog	c 792	34	63.0	396	24	ABS04795	Human genome-deriv
719	34.5	63.9	13401	22	AAK89558	Human digestive sy	c 793	34	63.0	396	24	ABS17309	Human genome-deriv
720	34.5	63.9	13403	22	AAK89557	Human digestive sy	c 794	34	63.0	400	22	AAK53444	Murine transport a
721	34.5	63.9	16373	22	ABA89934	Escherichia coli p	c 795	34	63.0	400	22	AAI92311	Human polynucleoti
722	34.5	63.9	23378	22	AAK89555	Human digestive sy	c 796	34	63.0	406	22	AAI91194	Human polynucleoti
723	34.5	63.9	23378	22	ABK42627	Genomic sequence #	c 797	34	63.0	410	13	AAQ28179	Human T lymphocyte
724	34	63.0	60	24	ABN45077	Human spliced tran	c 797	34	63.0	410	13	AAQ28179	Human T lymphocyte



c 798	34	63.0	414	22	AAI91450	Human polynucleoti	871	34	63.0	646	21	AAF14171	Aspergillus oryzae
c 799	34	63.0	416	20	AAAX41493	Human secreted pro	872	34	63.0	653	22	AAH22629	PPARGamma responsi
c 800	34	63.0	419	20	AAI86396	Human polynucleoti	c 873	34	63.0	678	22	AAK72649	Human immune/haema
c 801	34	63.0	420	20	AAI88067	Human polynucleoti	c 874	34	63.0	679	22	AAK72648	Human immune/haema
c 802	34	63.0	420	22	AAI15347	Chlamydia trachoma	c 875	34	63.0	682	21	AAA44881	Human secreted exp
c 803	34	63.0	420	22	AAI15452	Chlamydia trachoma	c 876	34	63.0	690	21	AAFI2953	Aspergillus oryzae
c 804	34	63.0	424	22	AAK64243	Human immune/haema	877	34	63.0	701	20	AAK35709	CDNA encoding a pr
c 805	34	63.0	425	24	ABN24721	Human OREFX polynuc	878	34	63.0	706	23	ABL09545	Drosophila melanog
c 806	34	63.0	430	22	AAI88192	Human polynucleoti	879	34	63.0	711	22	AAH80604	Receptor #92 parti
c 807	34	63.0	432	21	AAH51582	Human GGT5 related	c 880	34	63.0	725	22	AAI24778	Human breast cance
c 808	34	63.0	437	23	ABV58398	Human prostate exp	c 881	34	63.0	728	22	AAK58388	Human immune/haema
c 809	34	63.0	438	23	ABV16026	Human prostate exp	c 882	34	63.0	730	24	AAK55378	Human colon cancer
c 810	34	63.0	439	23	ABV45824	Human prostate exp	c 883	34	63.0	735	22	AAH67151	C glutamicum codin
c 811	34	63.0	441	22	AAH69814	Human cervical can	c 884	34	63.0	738	18	AAI95234	Macaque islet cell
c 812	34	63.0	442	22	AAI88366	Human polynucleoti	c 885	34	63.0	739	22	AAH70769	Human cervical can
c 813	34	63.0	445	20	AAI80544	Human polynucleoti	886	34	63.0	742	20	AAK40069	Human breast cance
c 814	34	63.0	450	20	AAK35708	CDNA encoding a pr	887	34	63.0	743	22	AAH98859	Colon cancer assoc
c 815	34	63.0	454	23	ABL25541	Drosophila melanog	c 888	34	63.0	745	24	ABT03062	Human EST-derived
c 816	34	63.0	456	22	AAH22633	PPARGamma responsi	c 889	34	63.0	748	24	ABK35977	Human breast speci
c 817	34	63.0	458	21	AAK35810	zebra mays DNA fragm	890	34	63.0	757	24	ABK73011	CDNA sequence #368
c 818	34	63.0	460	22	ABA58957	Human foetal liver	891	34	63.0	757	24	ABK73180	Bacillus lichenifo
c 819	34	63.0	460	22	ABA27816	Probe #6282 for ge	c 892	34	63.0	784	22	AAI23616	Human breast cance
c 820	34	63.0	460	22	AAK07119	Human brain expres	893	34	63.0	789	22	AAH22636	PPARGamma responsi
c 821	34	63.0	460	22	AAK32861	Human bone marrow	894	34	63.0	794	22	AAH71239	Human cervical can
c 822	34	63.0	460	22	AAI16237	Probe #6170 for ge	c 895	34	63.0	799	22	AAK87042	Human immune/haema
c 823	34	63.0	460	22	AAI38674	Probe #7360 used t	c 896	34	63.0	805	22	AAI14754	Human breast cance
c 824	34	63.0	460	22	ABS07665	Human genome-deriv	c 897	34	63.0	819	22	AAI94698	Human neuroblastom
c 825	34	63.0	465	24	ABK73833	Bacillus lichenifo	898	34	63.0	834	20	AAK23287	L. enzymogenes Lys
c 826	34	63.0	468	21	AAC47769	zebra mays DNA fragm	899	34	63.0	841	24	AAQ89989	Human prostate exp
c 827	34	63.0	474	21	ABA96840	Human transcriptio	900	34	63.0	857	22	AAI13498	Human breast cance
c 828	34	63.0	476	22	AAK63189	Human immune/haema	901	34	63.0	857	22	AAI13902	Human breast cance
c 829	34	63.0	477	22	ABA45811	Human breast cell	902	34	63.0	875	22	AAI05180	Human secreted pro
c 830	34	63.0	477	22	ABA56324	Human foetal liver	903	34	63.0	877	23	AAV15514	Human prostate exp
c 831	34	63.0	477	22	ABA25965	Probe #4431 for ge	904	34	63.0	883	22	AAI05377	Human secreted pro
c 832	34	63.0	477	22	AAK4504	Human brain expres	905	34	63.0	885	23	ABL08577	Drosophila melanog
c 833	34	63.0	477	22	AAK30005	Human bone marrow	c 906	34	63.0	888	22	AAK51735	Human polynucleoti
c 834	34	63.0	477	22	AAI14596	Probe #4529 for ge	c 907	34	63.0	890	22	AAH03553	Human CDNA clone (
c 835	34	63.0	477	22	AAI35968	Probe #4654 used t	c 908	34	63.0	899	22	AAH77057	Proliferative glom
c 836	34	63.0	477	22	AAI04410	Probe #4401 used t	909	34	63.0	906	22	AAH34459	Human colon cancer
c 837	34	63.0	477	24	ABS04577	Human genome-deriv	910	34	63.0	909	21	AAK58817	DNA encoding a hum
c 838	34	63.0	485	22	AAH69434	Human cervical can	911	34	63.0	913	22	AAI23770	Human breast cance
c 839	34	63.0	486	23	ABL24207	Drosophila melanog	c 912	34	63.0	914	23	ABL10435	Drosophila melanog
c 840	34	63.0	486	24	ABK74631	Bacillus lichenifo	c 913	34	63.0	916	20	AAK61468	DNA encoding a hum
c 841	34	63.0	491	22	AAH22632	PPARGamma responsi	c 914	34	63.0	932	18	AAI95235	Macaque islet cell
c 842	34	63.0	493	22	AAH22631	PPARGamma responsi	915	34	63.0	933	21	AAH77984	Nucleotide sequenc
c 843	34	63.0	502	24	ABO55346	Human ovarian anti	916	34	63.0	943	21	AAK58815	DNA encoding a hum
c 844	34	63.0	506	22	AAK70854	Human immune/haema	917	34	63.0	943	21	AAI291768	Human breast speci
c 845	34	63.0	507	18	AAK69733	Corn starch branch	918	34	63.0	958	22	AAK63163	Human purified sec
c 846	34	63.0	510	23	ABV06912	Human prostate exp	919	34	63.0	968	22	AAH72912	Human cervical can
c 847	34	63.0	524	24	ABN73636	Bovine embryonic g	920	34	63.0	968	22	AAH73229	Human cervical can
c 848	34	63.0	526	21	AAK38754	Wheat 1-deoxy-D-xy	921	34	63.0	978	21	AAA05580	Streptococcus pneu
c 849	34	63.0	526	22	AAH72312	Human cervical can	c 922	34	63.0	983	16	AAI03018	Endothelial cell p
c 850	34	63.0	536	24	ABN61040	Human cancer relat	c 923	34	63.0	983	24	ABK35835	CDNA sequence #226
c 851	34	63.0	543	22	AAH31366	Human secreted pro	c 924	34	63.0	985	21	AAH51301	Human GLC1 related
c 852	34	63.0	543	22	AAH31366	Human secreted pro	c 925	34	63.0	999	18	AAI95236	Macaque islet cell
c 853	34	63.0	543	23	ABV61361	Human secreted exp	c 926	34	63.0	1000	23	ABL30333	Drosophila melanog
c 854	34	63.0	546	21	AAK75593	Human prostate exp	c 927	34	63.0	1011	18	AAI95237	Macaque islet cell
c 855	34	63.0	546	24	ABN22399	Human OREFX polynuc	c 928	34	63.0	1029	21	AAI229084	Macaque islet cell
c 856	34	63.0	549	22	AAH71181	Human cervical can	c 929	34	63.0	1036	23	ABV30291	Human prostate exp
c 857	34	63.0	554	24	ABN65147	Human cancer relat	c 930	34	63.0	1036	20	AAI233474	Human prostate can
c 858	34	63.0	561	23	AAK68009	DNA encoding novel	c 931	34	63.0	1040	21	AAI21798	Human breast and o
c 859	34	63.0	573	22	AAK34153	Human CDNA encodin	c 932	34	63.0	1041	23	ABL11657	Drosophila melanog
c 860	34	63.0	574	21	AAI13617	Aspergillus oryzae	c 933	34	63.0	1047	21	AAK69513	Human secreted pro
c 861	34	63.0	587	24	ABK95322	Human prostate spe	c 934	34	63.0	1055	21	AAK79870	Human secreted pro
c 862	34	63.0	591	22	AAH90060	Human bone marrow	c 935	34	63.0	1061	22	AAK26068	Human CDNA encodin
c 863	34	63.0	594	22	AAK02054	DNA encoding molec	c 936	34	63.0	1064	23	ABK43693	DNA encoding novel
c 864	34	63.0	597	22	AAH22630	PPARGamma responsi	c 937	34	63.0	1064	24	ABT03063	Human breast speci
c 865	34	63.0	608	21	AAK75483	Nucleotide sequenc	c 938	34	63.0	1067	23	AAI71928	DNA encoding novel
c 866	34	63.0	618	16	AAI03017	Endothelial cell p	c 939	34	63.0	1067	22	AAK73053	DNA encoding novel
c 867	34	63.0	621	24	AAK62017	Porcine muscular p	c 940	34	63.0	1070	22	AAK51736	Human polynucleoti
c 868	34	63.0	634	23	ABV14408	Human prostate exp	c 941	34	63.0	1071	24	ABO86171	Novel human gene
c 869	34	63.0	639	20	AAI12165	Neisseria meningit	c 942	34	63.0	1071	21	AAK77283	Human ORFX ORF2838
c 870	34	63.0	639	21	AAK81374	N. meningitidis Me	c 943	34	63.0	1079	22	AAK52720	Human polynucleoti

c 944 34 63.0 1080 22 AAK52719 Human polynucleoti  
 c 945 34 63.0 1110 23 AAK29155 Pituitary somatost  
 c 946 34 63.0 1110 13 AB198025 Non-endogenous hum  
 c 947 34 63.0 1118 24 ABK33551 cDNA encoding huma  
 c 948 34 63.0 1121 22 AAK52316 Human polynucleoti  
 c 949 34 63.0 1121 22 AAK160271 Human polynucleoti  
 c 950 34 63.0 1130 17 AAF09250 Human somatostatin  
 c 951 34 63.0 1130 21 AAF13800 Aspergillus oryzae  
 c 952 34 63.0 1134 22 AAK51932 Human polynucleoti  
 c 953 34 63.0 1136 22 AAI58485 Human polynucleoti  
 c 954 34 63.0 1146 23 AB110197 Drosophila melanog  
 c 955 34 63.0 1161 22 AA544495 Botulism toxin hea  
 c 956 34 63.0 1185 23 AB111145 Drosophila melanog  
 c 957 34 63.0 1185 23 AAS24999 E. coli DNA for ce  
 c 958 34 63.0 1219 23 ABV23012 Human prostate exp  
 c 959 34 63.0 1219 23 ABV28848 Human prostate exp  
 c 960 34 63.0 1234 22 AAK76191 Human immune/haema  
 c 961 34 63.0 1234 22 AAK78001 Human immune/haema  
 c 962 34 63.0 1235 22 AAK78002 Human immune/haema  
 c 963 34 63.0 1235 22 AAK78003 Human immune/haema  
 c 964 34 63.0 1235 22 AAK78004 Human immune/haema  
 c 965 34 63.0 1242 21 AAK77765 Human cancer assoc  
 c 966 34 63.0 1244 14 AA045656 Murine somatostati  
 c 967 34 63.0 1245 23 ABV22314 Human prostate exp  
 c 968 34 63.0 1245 23 ABV28329 Human prostate exp  
 c 969 34 63.0 1248 23 AAS54858 Staphylococcus aur  
 c 970 34 63.0 1248 23 AAS55174 Staphylococcus aur  
 c 971 34 63.0 1249 19 AAV38987 cDNA encoding a hu  
 c 972 34 63.0 1249 21 AA58814 Human endometrium  
 c 973 34 63.0 1253 20 AAK42051 Human secreted pro  
 c 974 34 63.0 1255 22 ABA09371 Human prostate exp  
 c 975 34 63.0 1276 23 ABV22186 Human prostate exp  
 c 976 34 63.0 1276 23 ABV23043 Human prostate exp  
 c 977 34 63.0 1276 23 ABV28024 Human prostate exp  
 c 978 34 63.0 1287 23 ABV28879 Human prostate exp  
 c 979 34 63.0 1287 22 AAL26569 Human breast cance  
 c 980 34 63.0 1300 20 AAK39669 Renal cancer assoc  
 c 981 34 63.0 1309 7 AAN60856 Sequence encoding  
 c 982 34 63.0 1309 16 AAQ97288 Human interleukin-  
 c 983 34 63.0 1316 22 AAQ99480 Human protein enco  
 c 984 34 63.0 1320 20 AA207156 Human lung tumour  
 c 985 34 63.0 1320 21 AAK79067 Human lung tumour  
 c 986 34 63.0 1320 23 AAD23142 Human lung tumour  
 c 987 34 63.0 1328 20 AA207158 Human lung tumour  
 c 988 34 63.0 1328 21 AAK79069 Human lung tumour  
 c 989 34 63.0 1328 23 AAD23144 Human lung tumour  
 c 990 34 63.0 1331 20 AA207160 Human lung tumour  
 c 991 34 63.0 1331 21 AAK79072 Human lung tumour  
 c 992 34 63.0 1331 23 AAD23147 Human lung tumour  
 c 993 34 63.0 1333 20 AA207161 Human lung tumour  
 c 994 34 63.0 1333 21 AAK79073 Human lung tumour  
 c 995 34 63.0 1333 23 AAD23148 Human lung tumour  
 c 996 34 63.0 1347 19 AAV28624 Human aspartic pro  
 c 997 34 63.0 1347 20 AAK20605 Polynucleotide seq  
 c 998 34 63.0 1351 14 AA045655 Human somatostatin  
 c 999 34 63.0 1355 20 AA207164 Human lung tumour  
 1000 34 63.0 1355 21 AAK79076 Human lung tumour-

## ALIGNMENTS

RESULT 1  
 AAV75063  
 ID AAV75063 standard; DNA; 1019 BP.

AAV75063;

16-MAR-1999 (first entry)

Staphylococcus aureus contig SEQ ID #752.

Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

FH Key Location/Qualifiers  
 misc\_feature /\*tag= a

FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They  
 FT are included to maintain the nucleotide numbering  
 FT given in the specification for this DNA sequence".

XX EP786519-A2.

PD 30-JUL-1997.

PF 07-JAN-1997; 97EP-0100117.

PR 05-JAN-1996; 96US-0009861.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;  
 PI Rosen CA;

XX WPI; 1997-374922/35.

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines

Claim 1; Page 1646-1647; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.

XX SQ Sequence 1019 BP; 364 A; 194 C; 159 G; 240 T; 62 other;

Alignment Scores:

Pred. No.: 1.64 Length: 1019  
 Score: 54.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 18 Gaps: 0

US-10-008-355-25 (1-10) x AAV75063 (1-1019)

Oy 1 ThrGlyGlyAsnSerGlySerProValpHe 10

Db 642 ACTGGTGGTAATTCAGGTTACCTGATT 671

RESULT 2

AAQ27988

ID AAV75063 standard; DNA; 1558 BP.

XX

```

AC AAQ27988;
XX
DT 11-FEB-1993 (first entry)
XX
DE Protease from S. Aureus.
XX
KW Protease; PCR; amplify; Staphylococcus; ss.
XX
OS Staphylococcus aureus.
XX
XX
FH Key Location/Qualifiers
FT CDS 352..1425
FT sig_peptide /*tag= a
FT 352..555
FT mat_peptide /*tag= b
FT 556..1425
FT /*tag= c
XX
PN JP04211370-A.
XX
XX
PD 03-AUG-1992.
XX
PF 19-FEB-1991; 91JP-0024633.
XX
PR 20-FEB-1990; 90JP-0040398.
XX
XX (SHIO ) SHIONOGI & CO LTD.
XX
XX WPI; 1992-304938/37.
XX P-PSDB; AAR26842.
XX
XX Novel protease prep'd. using Bacillus or Saccharomyces host -
XX capable of cleaving peptide bond at carboxyl terminus of glutamic
XX acid residues in polypeptide(s)
XX
XX Disclosure; Page 13-15; 25pp; Japanese.
XX
XX The sequences given in AAQ27987-88 encode proteases which were isolated
XX from Staphylococcus aureus strains. The DNA sequences were isolated
XX by PCR using the primer sequences given in AAQ27960-86. The protease
XX specifically cleaves the peptide bond at the C-terminus of the
XX glutamic acid residue in polypeptide.
XX
XX Sequence 1586 BP; 579 A; 282 C; 221 G; 476 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 2.63 Length: 1558
XX Score: 54.00 Matches: 10
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 13 Gaps: 0
XX
XX US-10-008-355-25 (1-10) x AAQ27988 (1-1558)
XX
QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
XX |||||||
DB 1050 ACTGGTGGTAATTCAGGTTCCACCTGTATT 1079

RESULT 3
AAQ27987
ID AAQ27987 standard; DNA; 1586 BP.
XX
XX AAQ27987;
XX
XX 11-FEB-1993 (first entry)
XX
DE Protease from S. Aureus ATCC12600.
XX
XX Protease; PCR; amplify; Staphylococcus; ss.
XX
XX Staphylococcus aureus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..2139
FT /*tag= a
FT /*product= "Porphyromonas gingivalis DPP-7"
XX
XX Porphyromonas gingivalis.
XX
XX
OS Porphyromonas gingivalis.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..2139
FT /*tag= a
FT /*product= "Porphyromonas gingivalis DPP-7"
XX
XX WO200238742-A2.
XX
XX 16-MAY-2002.
XX
PD

```



XX WPI; 2002-381255/41.  
 DR P-PSDB; ABP38884.  
 XX  
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 PS Disclosure; SEQ ID 892; 267pp; English.  
 XX  
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 SQ Sequence 936 BP; 359 A; 128 C; 154 G; 295 T; 0 other;  
 XX

Alignment Scores:  
 Pred. No.: 12.1 Length: 936  
 Score: 49.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 90.74% Indels: 0  
 DB: 24 Gaps: 0

US-10-008-355-25 (1-10) x ABN91429 (1-936)  
 QY 2 GlyGlyAsnSerGlySerProValPhe 10  
 ID AAH54330/c  
 XX AAH54330 standard; DNA; 3189 BP.  
 XX  
 AC AAH54330;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3694.  
 XX  
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KW vaccination; endocarditis; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US030782.  
 XX  
 PR 09-NOV-1999; 99US-0164258.  
 XX  
 PR (GLAX ) GLAXO GROUP LTD.  
 PA  
 XX Kimmerly WJ;  
 PI  
 XX WPI; 2001-316495/33.  
 XX  
 DR Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 PT  
 XX  
 PS Claim 8; Page 1307-1308; 2188pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.

CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though the sequences are given in the disclosure for SEQ ID NO:4455 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 3189 BP; 1081 A; 596 C; 502 G; 1010 T; 0 other;  
 XX

Alignment Scores:  
 Pred. No.: 46.9 Length: 3189  
 Score: 49.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 90.74% Indels: 0  
 DB: 22 Gaps: 0

US-10-008-355-25 (1-10) x AAH54330 (1-3189)  
 QY 2 GlyGlyAsnSerGlySerProValPhe 10  
 ID AAL17622  
 DB 1147 GGTGGAACCTCTGGATCTCCAGTATT 1121  
 RESULT 8  
 AAL17622  
 ID AAL17622 standard; cDNA; 402 BP.  
 XX  
 AC AAL17622;  
 XX  
 DT 07-DEC-2001 (first entry)  
 XX  
 DE Human breast cancer expressed polynucleotide 10079.  
 XX  
 KW Human; breast cancer; cell marker; cytostatic; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151628-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 10-JAN-2001; 2001WO-US00798.  
 XX  
 PR 14-JAN-2000; 2000US-0176077.  
 PR 14-MAR-2000; 2000US-0189167.  
 PR 24-MAR-2000; 2000US-0192099.  
 PR 29-MAR-2000; 2000US-0193480.  
 PR 15-MAY-2000; 2000US-0205230.  
 PR 09-JUN-2000; 2000US-0211315.  
 PR 25-JUL-2000; 2000US-0220534.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Lillie J, Xu Y, Wang Y, Steinmann K;  
 XX WPI; 2001-451856/48.  
 DR  
 XX  
 PT New peptide useful as a marker for the diagnosis of breast cancer  
 PS Claim 1; Page 1799; 3695pp; English.  
 XX

CC The invention relates to human breast cancer expressed polynucleotides  
CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
CC afflicted with breast cancer by examining the correlation between the  
CC expression of certain markers and the cancerous state of breast cells.  
CC The polynucleotides and encoded polypeptides are potential markers for  
CC detecting, diagnosing, monitoring, characterising treating and  
CC potentially preventing breast cancer. The polynucleotides and encoded  
CC polypeptides are also useful for isolating compounds with cytostatic  
CC activity.

SQ Sequence 402 BP; 101 A; 98 C; 104 G; 99 T; 0 other;

#### Alignment Scores:

Pred. No.:	38.4	Length:	402
Score:	44.00	Matches:	7
Percent Similarity:	90.00%	Conservative:	2
Best Local Similarity:	70.00%	Mismatches:	1
Query Match:	81.48%	Indels:	0
DB:	22	Gaps:	0

US-10-008-355-25 (1-10) x AAL17622 (1-402)

Oy 1 ThrGlyGlyAsnSerGlySerProValphe 10

Db 294 ACCGGGGGAGTTCGGGAGAACCCATTTT 323

#### RESULT 9

AAL00017

ID AAL00017 standard; cDNA; 423 BP.

XX AC AAL00017;

XX XX 21-NOV-2001 (first entry)

XX Human reproductive system related antigen cDNA SEQ ID NO: 18.

XX Human; reproductive system related antigen; reproductive system disorder;  
XX cancer; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01339.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180828.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.

AAX20213

ID AAX20213 standard; DNA; 758 BP.

XX AC

XX AC

XX AC AAX20213;

XX DT 20-APR-1999 (first entry)

XX DE Enterococcus faecalis EF110 gene fragment.

XX KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;  
detection; attenuation; antigenic; ss.

XX OS Enterococcus faecalis.

XX PN WO9850554-A2.

XX PD 12-NOV-1998.

XX PF 04-MAY-1998; 98WO-US08959.

XX PR 14-NOV-1997; 97US-0066009.

XX PR 06-MAY-1997; 97US-0044031.

XX PR 16-MAY-1997; 97US-0046655.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PA Bailey C, Choi GH, Hromockyj A, Kunsch CA;

XX PI WPI; 1999-070095/06.

XX DR P-PADB; AAY00223.

XX PT New isolated Enterococcus faecalis polynucleotides - used to develop  
products for the detection of Enterococcus and for use in vaccines  
for prevention or attenuation of Enterococcus infection

XX PS Claim 1; Page 212-213; 301pp; English.

XX XX The present sequence encodes an antigenic polypeptide fragment  
isolated from Enterococcus faecalis. The present invention describes  
genes, proteins and antigenic polypeptides isolated from E. faecalis.  
The proteins can be used in vaccines for preventing or attenuating an  
infection caused by a member of the Enterococcus genus in an animal.  
They can also be used for detecting Enterococcus antibodies in a sample  
The nucleotide sequences can be used for detecting Enterococcus nucleic  
acids. Products from the present invention can also be used for  
screening compounds to identify agonists and antagonists of E. faecalis  
protein activity.

XX SQ Sequence 758 BP; 248 A; 143 C; 160 G; 207 T; 0 other;

Alignment Scores:

Pred. No.: 77.5 Length: 758

Score: 44.00 Matches: 7

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 81.48% Indels: 0

DB: 20 Gaps: 0

US-10-008-355-25 (1-10) x AAX20213 (1-758)

QY 1 ThrGlyGlyAsnSerGlySerProValphe 10

||||||| ||||||| |:::|

Db 575 ACCGGCGGTCAATCTGGTCACCAATCTAT 604

RESULT l1

ABN98198

ID ID ABN98198 standard; DNA; 758 BP.

XX AC

XX AC ABN98198;

XX XX

XX DT 05-AUG-2002 (first entry)

XX DE E faecalis EF110 gene fragment

XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;  
KW gene; ds.  
OS Enterococcus faecalis.  
XX US2002045737-A1.  
PN 18-APR-2002.  
XX 04-MAY-1998; 98US-0071035.  
PF 04-MAY-1998; 98US-0071035.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;  
PI WPI; 2002-425450/45.  
XX P-PSDB; ABP43442.  
DR New genes and polypeptides from Enterococcus faecalis, useful as  
XX vaccines for preventing, treating or attenuating an infection caused by  
PT a member of the Enterococcus genus in an animal, particularly E.  
PT faecalis.  
XX Claim 1; Page 197-198; 255pp; English.  
PS The present invention provides the protein and coding sequences of a  
XX number of polypeptides from Enterococcus faecalis. The proteins can be  
CC used as vaccines for preventing or attenuating an infection caused by a  
CC member of the Enterococcus genus in an animal, particularly E. faecalis.  
CC The polynucleotide is also useful for preventing or treating E. faecalis  
CC infection. The present sequence is a coding sequence of the invention.  
XX Sequence 758 BP; 248 A; 143 C; 160 G; 207 T; 0 other;  
SQ Alignment Scores:  
Pred. No.: 77.5 Length: 758  
Score: 44.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 81.48% Indels: 0  
DB: 24 Gaps: 0  
US-10-008-355-25 (1-10) x ABN98198 (1-758)  
Qy 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
Db 575 ACCGGCGTCAATCTGGTTCACCAATCTAT 604  
RESULT 12  
AAX20212  
ID AAX20212 standard; DNA; 888 BP.  
XX AAX20212;  
XX 20-APR-1999 (first entry)  
XX Enterococcus faecalis gene EF110.  
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;  
KW detection; attenuation; antigenic; ds.  
XX Enterococcus faecalis.  
XX WO98050554-A2.  
PN 12-NOV-1998.  
XX 04-MAY-1998; 98WO-US089959.  
PF 14-NOV-1997; 97US-0066009.  
PR

PR 06-MAY-1997; 97US-0044031.  
XX 16-MAY-1997; 97US-0046655.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Bailey C, Choi GH, Hromockyj A, Kunsch CA;  
PI WPI; 1999-070095/06.  
XX P-PSDB; AAY00222.  
DR New isolated Enterococcus faecalis polynucleotides - used to develop  
XX products for the detection of Enterococcus and for use in vaccines  
XX for prevention or attenuation of Enterococcus infection  
XX Claim 1; Page 212; 301pp; English.  
XX The present sequence represents a gene isolated from  
CC Enterococcus faecalis. The present invention describes genes, proteins  
CC and antigenic polypeptides isolated from E. faecalis. The proteins can  
CC be used in vaccines for preventing or attenuating an infection caused  
CC by a member of the Enterococcus genus in an animal. They can also be  
CC used for detecting Enterococcus antibodies in a sample. The nucleotide  
CC sequences can be used for detecting Enterococcus nucleic acids.  
CC Products from the present invention can also be used for screening  
CC compounds to identify agonists and antagonists of E. faecalis protein  
CC activity.  
XX Sequence 888 BP; 230 A; 159 C; 186 G; 253 T; 0 other;  
SQ Alignment Scores:  
Pred. No.: 92.3 Length: 888  
Score: 44.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 81.48% Indels: 0  
DB: 20 Gaps: 0  
US-10-008-355-25 (1-10) x AAX20212 (1-888)  
Qy 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
Db 703 ACCGGCGTCAATCTGGTTCACCAATCTAT 732  
RESULT 13  
ABN98197  
ID ABN98197 standard; DNA; 888 BP.  
XX ABN98197;  
XX 05-AUG-2002 (first entry)  
XX E faecalis EF110 gene.  
XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;  
KW gene; ds.  
XX Enterococcus faecalis.  
XX US2002045737-A1.  
PN 18-APR-2002.  
XX 04-MAY-1998; 98US-0071035.  
PF 04-MAY-1998; 98US-0071035.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;  
PI WPI; 2002-425450/45.  
XX P-PSDB; ABP43441.  
DR



PT New genes and polypeptides from *Enterococcus faecalis*, useful as  
 PT vaccines for preventing, treating or attenuating an infection caused by  
 PT a member of the *Enterococcus* genus in an animal, particularly *E.*  
 PT *faecalis* -

PS Claim 1; Page 197; 255pp; English.

XX The present invention provides the protein and coding sequences of a  
 CC number of polypeptides from *Enterococcus faecalis*. The proteins can be  
 CC used as vaccines for preventing or attenuating an infection caused by a  
 CC member of the *Enterococcus* genus in an animal, particularly *E. faecalis*.  
 CC The polynucleotide is also useful for preventing or treating *E. faecalis*  
 CC infection. The present sequence is a coding sequence of the invention.

XX SQ Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;

Alignment Scores:  
 Pred. No.: 92.3 Length: 888  
 Score: 44.00 Matches: 7  
 Percent Similarity: 90.00% Conservative: 2  
 Best Local Similarity: 70.00% Mismatches: 1  
 Query Match: 81.48% Indels: 0  
 DB: 24 Gaps: 0

US-10-008-355-25 (1-10) x ABN98197 (1-888)

QY 1 ThrGlyAsnSerGlySerProValphe 10

Db 703 ACCGGCGGTCAATCTGTTCCACCAATCTAT 732

RESULT 14

AAV80633

ID AAV80633 standard; cDNA; 2421 BP.

XX AC AAV80633;

XX 01-MAR-1999 (first entry)

XX Kidney injury associated molecule HW096 cDNA clone.

XX Kidney injury associated molecule; kidney injury related molecule;  
 KW KIM; tissue growth promotion; regeneration; renal condition;  
 KW acute renal failure; acute nephritis; tumour; ds.

XX OS Rattus sp.

XX FH Key Location/Qualifiers

FT CDS 295..1305

FT /\*tag= a

FT /product= "kidney injury associated molecule"

XX WO9853071-A1.

XX PN 26-NOV-1998.

XX PF 22-MAY-1998; 98WO-US10547.

XX PR 23-MAY-1997; 97US-0047491.

XX PR 23-MAY-1997; 97US-0047490.

XX PA (BIOJ ) BIOGEN INC.

XX PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;

XX DR WPI: 1999-045312/04.

XX DR P-PSDB; AAW86338.

XX Kidney injury-associated molecule, KIM, polypeptides - upregulated  
 PT in injured or regenerating tissues, useful to promote tissue growth  
 PT and regeneration, especially to treat renal conditions

PS Claim 9; Page 186-188; 213pp; English.

XX

CC The present sequence represents a kidney injury associated molecule  
 CC (KIM) cDNA clone. KIM proteins can be administered therapeutically  
 CC by expressing KIM encoding polynucleotides, to promote growth and/or  
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins  
 CC are upregulated in injured or regenerating (especially renal) tissues.  
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used  
 CC therapeutically, e.g. these or the KIM proteins may be included with an  
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/  
 CC prophylaxis of conditions associated with dysfunction/dysregulation of  
 CC KIM genes or proteins, especially renal diseases or impairments of renal  
 CC function in humans (e.g. acute renal failure, acute nephritis). The  
 CC polynucleotides can be used to produce antisense sequences which, when  
 CC internalised into cells, can disrupt expression of a cellular KIM gene,  
 CC also useful in therapy (e.g. to block the growth of tumours dependent on  
 CC KIM for growth) or compositions. The proteins and polynucleotides are  
 CC useful diagnostically e.g. to detect and quantify renal injury/disease  
 CC (indicative of increased risk, or presence of, renal injury or impaired  
 CC function), or abnormal responses to tissue injury (indicative of  
 CC increased risk, or presence of, an autoimmune response or abnormal  
 CC tissue growth arising from/affecting renal tissue). The proteins can  
 CC also be used to locate KIM-producing cells (especially specific loci,  
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours  
 CC arising from/affecting renal tissue), by contacting cells with an  
 CC imageable KIM-binding reagent and imaging reagent accumulation.

XX SQ Sequence 2421 BP; 785 A; 417 C; 533 G; 686 T; 0 other;

Alignment Scores:

Pred. No.: 280 Length: 2421  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 20 Gaps: 0

US-10-008-355-25 (1-10) x AAV80633 (1-2421)

QY 2 GlyGlyAsnSerGlySerProValphe 10

Db 1087 GGTGGTCTTCTGGATCCCGAGTATTT 1113

RESULT 15

AAV13198/c

ID AAV13198 standard; DNA; 4951 BP.

XX AC AAV13198;

XX 19-MAR-1999 (first entry)

XX DE *Enterococcus faecalis* genome contig SEQ ID NO:261.

XX KW *Enterococcus faecalis*; contig; detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.

XX OS *Enterococcus faecalis*.

XX PN WO9850555-A2;

XX PD 12-NOV-1998.

XX PF 04-MAY-1998; 98WO-US08985.

XX PR 14-NOV-1997; 97US-0066009.

XX PR 06-MAY-1997; 97US-0044031.

XX PR 16-MAY-1997; 97US-0046655.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Dillion PJ, Kunsch CA;

XX DR WPI: 1999-045171/04.

XX PT New isolated *Enterococcus faecalis* polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
PT infection.  
XX  
XX Claim 1; Page 1248-1251; 2084pp; English.  
XX  
XX A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC AAX12938 to AAX13919 represent these nucleotide sequences which are  
CC primary nucleotide sequences, also known as contigs. The computer-based  
CC system can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
CC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.  
XX  
XX Sequence 4951 BP; 1563 A; 981 C; 735 G; 1665 T; 7 other;  
SQ  
Alignment Scores:  
Pred. No.: 618 Length: 4951  
Score: 44.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 81.48% Indels: 0  
DB: 20 Gaps: 0  
US-10-008-355-25 (1-10) x AAX13198 (1-4951)  
Qy 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
Db 695 ACCGGCGTCAATCTGGTTACCAATCTAT 666  
RESULT 16  
ABL30036/c  
ID ABL30036 standard; DNA: 17391 BP.  
XX  
XX ABL30036;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DT  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 41581.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
PR  
XX  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
PT  
XX  
XX Claim 1; SEQ ID NO 41581; 21pp + Sequence Listing; English.  
XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 17391 BP; 4874 A; 3695 C; 3785 G; 5037 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 2.48e+03 Length: 17391  
Score: 44.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 81.48% Indels: 0  
DB: 23 Gaps: 0  
US-10-008-355-25 (1-10) x ABL30036 (1-17391)  
Qy 2 GlyGlyAsnSerGlySerProValPhe 10  
Db 8234 GGGGGGAGTTCGGGGAGCCAGTATT 8208  
RESULT 17  
ABK73036  
ID ABK73036 standard; DNA: 534 BP.  
XX  
XX ABK73036;  
AC  
XX  
XX 13-AUG-2002 (first entry)  
DT  
DE Bacillus licheniformis genomic sequence tag (GST) #327.  
XX  
XX Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
XX Bacillus licheniformis.  
OS  
XX  
XX WO200229113-A2.  
PN  
XX  
XX 11-APR-2002.  
PD  
XX  
XX 05-OCT-2001; 2001WO-US31437.  
PF  
XX  
XX 06-OCT-2000; 2000US-0680598.  
PR  
XX  
XX 27-MAR-2001; 2001US-279526P.  
PR  
XX  
XX (NOVO ) NOVOZYMES BIOTECH INC.  
PA  
XX  
XX (NOVO ) NOVOZYMES AS.  
PA  
XX  
XX Berka R, Clausen IG;  
PI  
XX  
XX WPI; 2002-416684/44.  
DR  
XX  
XX Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second  
PT Bacillus cells, by using substrate containing Bacillus genomic  
PT sequenced tag array.  
PT  
XX  
XX Claim 4; SEQ ID NO 327; 200pp; English.  
PS  
XX  
XX The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of

a spot in the array. The method is useful for measuring the expression of genes in a first Bacillus cell relative to expression of the same genes in one or more second Bacillus cells. The method is useful for monitoring global expression of several genes from a Bacillus cell, discovering new genes, identifying possible functions of unknown open reading frames and monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which Bacillus cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

SQ Sequence 534 BP; 158 A; 144 C; 128 G; 104 T; 0 other;

Alignment Scores:	79.9	Length:	534
Pred. No.:	43.00	Matches:	8
Score:	88.89%	Conservative:	0
Percent Similarity:	88.89%	Mismatches:	1
Best Local Similarity:	88.89%	Indels:	0
Query Match:	79.63%	Gaps:	0
DB:	24		

US-10-008-355-25 (1-10) x ABK73036 (1-534)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10

RESULT 18

AAD02999  
ID AAD02999 standard; DNA: 948 BP.

DT 31-MAY-2001 (first entry)

DE Bacillus licheniformis (BLC) RP-II protease DNA.

Residual protease II; RP-II; additive; cleaning composition; detergent; ds.

XX Bacillus Licheniformis.

Key	Location/Qualifiers
CDS	1..948
FT	/*tag= a
FT	/product= "Bacillus licheniformis RP-II protease"
FT	/note= "CDS does not include both start and stop
FT	codon"
FT	/partial
FT	1..93
FT	/*tag= b
FT	94..282
FT	/*tag= c
FT	/note= "Pro sequence of RP-II protease"
FT	283..948
mat_peptide	/*tag= d
FT	/product= "Bacillus licheniformis mature RP-II pr

AA  
PN  
WO200116285-A2XX  
PD  
08-MAR-2001.

31-AUG-2000: 2000WO-DK00476.

XX 31-AUG-1999: 99PK-0001212

PR 31-AUG-1999; 99DK-0001212;  
PR 20-OCT-1999; 99DK-0001500;

XX	( NOVO ) NOVOZYMES AS.	
PA		
XX		
PI	Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;	
PI	Flensted Lassen S;	
XX		
XX	WPI: 2001-226680/23.	
DR	P-PSDB; AAE00011.	
DR		
XX		
PT	Novel RP-II type protease and its variants useful as constituents in	
PT	detergent compositions, additives and cleaning compositions -	
XX		
XX		
PS	Claim 26; Page 99-101; 132pp; English.	
XX		
CC	The present sequence is Bacillus licheniformis (BLC) RP-II (Residual	
CC	protease II) DNA. RP-II protease is useful as a constituent in addit	
CC	detergent compositions and other cleaning compositions, optionally i	
CC	combination with other enzymes such as proteases, lipases, cellulase	
CC	amylases, peroxidases or oxidases. The variants of RP-II have improv	
CC	properties such as substrate specificities, catalytic rate, stabilit	
CC	especially towards the action of proteolytic enzymes and improved	
CC	resistance towards peroxidase.	

Sequence 948 BP; 287 A; 241 C; 218 G; 202 T; 0 other;

Alignment Scores:			
Pred. No.:	.	151	Length: -
Score:	.	43.00	Matches: 8
Percent Similarity:	88.89%		Conservative: 0
Best Local Similarity:	88.89%		Mismatches: 1
Query Match:	79.63%		Indels: 0
DB:	22		Gaps: 0

US-10-008-355-25 (1-10) x AAD02999 (1-948)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10

RESULT 19

AAQ24382  
ID AAQ24382 standard; DNA; 1448 BP.

XX  
DT 30-OCT-1992 : (first entry)XX  
DE  
Protease Blase

XX  
KW  
KW  
KW  
KW

XX  
OS Bacillus licheniformis.

key	Location/Qualifiers
CDS	323
	/tag= a
	/label= Blase_Protease_gene
sig_peptide	323..604
	/tag= b
mat_peptide	605..1270
	/tag= c

XX  
PN EP482879-A.XX  
 DD 29-APR-1992

XX  
PF 22-JCT-1991. 91EP-0309737

XX  
2A-OCT-1990.  
90.TP-0288110XX  
XX  
PA (SHTO) SHIONOGI & CO. LTD.

PA (SHIO ) SHIONOGI & CO LTD.  
PA (SHIO ) SHIONOGI SETYAKU KK.

XX Fujiwara T, Matsumoto K, Nakamura E, Shin M, Tamaki M;  
PI Teraoka H, Tsuzuki H, Yoshida N;  
XX WPI: 1992-142934/18.  
DR P-PSDB; AAR23730.  
XX Protease derived from Bacillus licheniformis - useful for e.g.  
PT seqq. fusion protein components by cleaving glutamic acid  
PT residues  
XX Disclosure: Page 19; 32pp; English.  
PS  
XX The sequence give encodes a novel protease derived from Bacillus  
CC licheniformis designated Blase. Blase specifically cleaves the  
CC peptide bond at the carboxyl terminal of glutamic acid residues in the  
CC amino acid sequence of polypeptides. The protease has an optimal pH  
CC of approx 8.0 with a stable pH range of 6.5-8.5. Blase is useful for  
CC cleaving proteins to enable protein structural analysis or for  
CC separating components of fusion proteins linked by a glutamic acid  
CC residue.  
XX  
SQ Sequence 1448 BP; 431 A; 360 C; 294 G; 363 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 241 Length: 1448  
Score: 43.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 79.63% Indels: 0  
DB: 13 Gaps: 0  
  
US-10-008-355-25 (1-10) x AQA24382 (1-1448)  
  
Qy 2 GlyGlyAsnSerGlySerProValpHe 10  
Db 1094 GGAGGACAAAGCGGTTACCGGTATTC 1120  
  
RESULT 20  
AAF79971  
ID AAF79971 standard; DNA; 152 BP.  
XX  
AC AAF79971;  
XX  
DT 11-JUN-2001 (first entry)  
XX  
DE Nucleotide sequence of a human genetic marker for toxicity.  
XX  
KW Genetic marker; toxicity; cellular signalling pathway; polymorphism; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200120029-A2.  
XX  
PD 22-MAR-2001.  
XX  
PF 12-SEP-2000; 2000WO-FR02503.  
XX  
PR 13-SEP-1999; 99FR-0011405.  
XX  
PA (EXON-) EXONHIT THERAPEUTICS SA.  
XX  
PI Tocque B, Bracco L, Schweighoffer F;  
XX  
DR WPI: 2001-244821/25.  
XX  
XX Analysing the toxic potential of test compounds, for e.g. screening  
PT for toxic effects in potential pharmaceuticals, comprises analysing  
PT hybridisation patterns of treated cells  
XX  
PS Claim 35; Page 57; 68pp; French.  
XX  
CC AAF79967-AAF80003 represents genetic markers of toxicity. The

CC specification describes a method for analysing the toxic potential  
CC of a test compound. The method comprises hybridising nucleic acids  
CC from cells treated with the test compound and the present markers.  
CC These markers correspond to genetic events characteristic of  
CC deregulation of cellular signalling pathways. The method is used to  
CC identify the toxic potential of compounds (particularly human or  
CC veterinary pharmaceuticals or plant protection agents) and to evaluate  
CC the response and/or sensitivity of subjects to a particular compound,  
CC from the presence of polymorphisms or other mutations in particular  
CC genes.  
XX  
SQ Sequence 152 BP; 28 A; 47 C; 35 G; 42 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 30.2 Length: 152  
Score: 42.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 3  
Best Local Similarity: 66.67% Mismatches: 0  
Query Match: 77.78% Indels: 0  
DB: 22 Gaps: 0  
  
US-10-008-355-25 (1-10) x AAF79971 (1-152)  
  
Qy 2 GlyGlyAsnSerGlySerProValpHe 10  
Db 26 GGGGGGAACCTCAGGCAATCCGATTAC 52  
  
RESULT 21  
ABV56742/C  
ID ABV56742 standard; cDNA; 473 BP.  
XX  
AC ABV56742;  
XX  
DT 17-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 56733.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI: 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
PS Claim 1; Page 10937; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate



PR 29-JUL-1999; 99US-0146453.  
PR 28-OCT-1999; 99US-0162288.  
XX (GEST ) GENSET.  
XX Cohen D, Blumenfeld M, Chumakov I, Bougueleret L, Bihain B;  
PI Essioux L;  
XX WPI; 2000-619082/59.  
XX Polynucleotides comprising sequences from sbg1 and g35018 biallelic  
PT markers are used for genotyping and detecting schizophrenia or bipolar  
PT disorder and predisposition to these disorders -  
XX Claim 1; Page 409-493; 737pp: English.  
XX AAH51601 represents a human genomic nucleotide sequence comprising sbg1,  
CC g3465, sbg2, g35017 and g35018 nucleic acid sequences located on the  
CC human chromosome 13q31-q33 locus. The nucleotide sequences contain  
CC biallelic markers and polymorphisms. Sequences AAH51602 - AAH51626 and  
CC AAB62907 - AAB62915 represent cDNA human sbg1 cDNA sequences and protein  
CC products. AAH51627 - AAH51631 and AAB62916 - AAB62918 represent g35018  
CC cDNA sequences and protein products. Primers AAH51632 - AAH51699 are used  
CC to isolate sbg1 cDNAs, while sbg1 exons from different primates are  
CC represented by sequences AAH51642 - AAH51699. Nucleotide sequences of  
CC amplicons which comprise biallelic markers located on the chromosome  
CC 13q31-q33 locus are represented in AAH51700 - AAH51817. Biallelic markers  
CC are represented in the sequences by degenerate/undefined base codes. PCR  
CC primers AAH51818 and AAH51819 are used in the isolation of sequences of  
CC the invention. The biallelic marker containing nucleotide sequences are  
CC used to determine the identity of the nucleotide at a biallelic marker in  
CC a sample DNA sequence. The nucleotide sequences may be labelled and used  
CC for genotyping by determining the identity of a nucleotide at a region  
CC D-related biallelic marker in a biological sample from single or multiple  
CC subjects. By determining the frequency of a biallelic marker in a  
CC population an association between a genotype and a trait, a haplotype and  
CC a trait and a phenotype and a trait can be detected. The sequences can be  
CC used to determine a predisposition to or early onset of schizophrenia or  
CC bipolar disorder or a beneficial response to or side effects related to  
CC treatment against schizophrenia or bipolar disorder.  
XX Sequence 319608 BP; 101600 A; 56677 C; 58335 G; 102722 T; 274 other;  
  
Alignment Scores:  
Pred. No.: 1.41e+05 Length: 319608  
Score: 42.00 Matches: 7  
Percent Similarity: 88.89% Conservative: 1  
Best Local Similarity: 77.78% Mismatches: 1  
Query Match: 77.78% Indels: 0  
DB: 21 Gaps: 0  
  
US-10-008-355-25 (1-10) x AAH51601 (1-319608)  
  
QY 2 GlyGlyAsnSerGlySerProValpHe 10  
||||||| |||:|||||||  
Db 27135 GGTGGGAATGTTGGGAACCCGTGTGT 27109  
  
RESULT 24  
AA509301/c  
ID AA509301 standard; DNA; 319608 BP.  
XX AC AA509301;  
XX 26-SEP-2001 (first entry)  
XX Human schizophrenia associated gene g35030 and biallelic markers A1-A71.  
DE Human; g35030; biallelic marker; A1-A71; chromosome 13q31-q33;  
XX Human; g35030; biallelic marker; A1-A71; chromosome 13q31-q33;  
KW schizophrenia; bipolar disorder; ds.  
XX Homo sapiens.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FH

FT primer\_bind 7938..7958  
FT /\*tag= a  
FT /note= "Binds primer 99-27943.rp"  
FT 8297..8315  
FT /\*tag= b  
FT /note= "Binds primer 99-27943-150.mis"  
FT 8304..8328  
FT /\*tag= c  
FT /bound\_moiety= Probe\_99-27943-150  
FT 8316  
FT /\*tag= d  
FT /note= "Biallelic marker A1"  
FT complement (8317..8335)  
FT /\*tag= e  
FT /note= "Binds primer 99-27943-150.mis complement"  
FT complement (8446..8465)  
FT /\*tag= f  
FT /note= "Binds primer 99-27943.pu complement"  
FT 21365..21385  
FT /\*tag= g  
FT /note= "Binds primer 99-27935.rp"  
FT 21653..21671  
FT /\*tag= h  
FT /note= "Binds primer 99-27935-193.mis"  
FT 21660..21684  
FT /\*tag= i  
FT /bound\_moiety= Probe\_99-27935-193  
FT 21672  
FT /\*tag= j  
FT /note= "Biallelic marker A2"  
FT complement (21673..21691)  
FT /\*tag= k  
FT /note= "Binds primer 99-27935-193.mis complement"  
FT complement (21845..21864)  
FT /\*tag= l  
FT /note= "Binds primer 99-27935.pu complement"  
FT 65463..65471  
FT /\*tag= m  
FT /note= "Binds primer 8-128.pu"  
FT 65466..65484  
FT /\*tag= n  
FT /note= "Binds primer 8-128-33.mis"  
FT 65473..65497  
FT /\*tag= o  
FT /bound\_moiety= Probe\_8-128-33  
FT 65485  
FT /\*tag= p  
FT /note= "Biallelic marker A3"  
FT complement (65486..65504)  
FT /\*tag= q  
FT /note= "Binds primer 8-128-33.mis complement"  
FT complement (65856..65874)  
FT /\*tag= r  
FT /note= "Binds primer 8-128.rp complement"  
FT 95034..95053  
FT /\*tag= s  
FT /note= "Binds primer 99-31960.pu"  
FT 95377..95395  
FT /\*tag= t  
FT /note= "Binds primer 99-31960-363.mis"  
FT 95384..95408  
FT /\*tag= u  
FT /bound\_moiety= Probe\_99-31960-363  
FT 95396  
FT /\*tag= v  
FT /note= "Biallelic marker A4"  
FT complement (95397..95415)  
FT /\*tag= w  
FT /note= "Binds primer 99-31960-363.mis complement"  
FT complement (95543..95563)  
FT /\*tag= x  
FT /note= "Binds primer 99-31960.rp complement"  
FT 107022..107040

```
FT FT primer_bind /tag= Y /note= "Binds primer 99-24656.pu"  
FT FT 107262..107280  
FT FT /tag= Z  
FT FT /note= "Binds primer 99-24656-260.mis"  
FT FT 107269..107293  
FT FT /tag= aa  
FT FT /bound_moiety= Probe_99-24656-260  
FT FT 107281  
FT FT /tag= ab  
FT FT /note= "Biallelic marker A5"  
FT FT complement (107282...107300)  
FT FT /tag= ac  
FT FT /note= "Binds primer 99-24656-260.mis complement"  
FT FT complement (107495...107513)  
FT FT /tag= ad  
FT FT /note= "Binds primer 99-24656.rp complement"  
FT FT 160279..160298  
FT FT /tag= ae  
FT FT /note= "Binds primer 99-24639.rp"  
FT FT 160621..160639  
FT FT /tag= af  
FT FT /note= "Binds primer 99-24639-163.mis"  
FT FT 160628..160652  
FT FT /tag= ag  
FT FT /bound_moiety= Probe_99-24639-163  
FT FT 160640  
FT FT /tag= ah  
FT FT /note= "Biallelic marker A6"  
FT FT complement (160641...160659)  
FT FT /tag= ai  
FT FT /note= "Binds primer 99-24639-163.mis complement"  
FT FT 160770..160787  
FT FT /tag= aj  
FT FT /note= "Binds primer 99-24634.pu"  
FT FT complement (160785...160802)  
FT FT /tag= ak  
FT FT /note= "Binds primer 99-24639.pu complement"  
FT FT 160857..160875  
FT FT /tag= al  
FT FT /note= "Binds primer 99-24634-108.mis"  
FT FT 160864..160888  
FT FT /tag= am  
FT FT /bound_moiety= Probe_99-24634-108  
FT FT 160876  
FT FT /tag= an  
FT FT /note= "Biallelic marker A7"  
FT FT complement (160877...160895)  
FT FT /tag= ao  
FT FT /note= "Binds primer 99-24634-108.mis complement"  
FT FT complement (161240...161257)  
FT FT /tag= ap  
FT FT /note= "Binds primer 99-24634.rp complement"  
FT FT 168813..168830  
FT FT /tag= aq  
FT FT /note= "Binds primer 99-7652.pu"  
FT FT 168955..168973  
FT FT /tag= ar  
FT FT /note= "Binds primer 99-7652-162.mis"  
FT FT 168962..168986  
FT FT /tag= as  
FT FT /bound_moiety= Probe_99-7652-162  
FT FT 168974  
FT FT /tag= at  
FT FT /note= "Biallelic marker A8"  
FT FT complement (168975...168993)  
FT FT /tag= au  
FT FT /note= "Binds primer 99-7652-162.mis complement"  
FT FT complement (169331...169351)  
FT FT /tag= av  
FT FT /note= "Binds primer 99-7652.rp complement"  
FT FT 170666..170686  
FT FT /tag= aw
```

```
FT FT primer_bind /note= "Binds primer 99-16100.pu"  
FT FT 170791..170809  
FT FT /tag= ax  
FT FT /note= "Binds primer 99-16100-147.mis"  
FT FT 170798..170822  
FT FT /tag= ay  
FT FT /bound_moiety= Probe_99-16100-147  
FT FT 170810  
FT FT /tag= az  
FT FT /note= "Biallelic marker A9"  
FT FT complement (170811...170829)  
FT FT /tag= ba  
FT FT /note= "Binds primer 99-16100-147.mis complement"  
FT FT complement (171153...171173)  
FT FT /tag= bb  
FT FT /note= "Binds primer 99-16100.rp complement"  
FT FT 173065..173085  
FT FT /tag= bc  
FT FT /note= "Binds primer 99-5862.rp"  
FT FT 173339..173357  
FT FT /tag= bd  
FT FT /note= "Binds primer 99-5862-167.mis"  
FT FT 173346..173370  
FT FT /tag= be  
FT FT /bound_moiety= Probe_99-5862-167  
FT FT 173358  
FT FT /tag= bf  
FT FT /note= "Biallelic marker A10"  
FT FT complement (173359...173377)  
FT FT /tag= bg  
FT FT /note= "Binds primer 99-5862-167.mis complement"  
FT FT complement (173495...173514)  
FT FT /tag= bh  
FT FT /note= "Binds primer 99-5862.pu complement"  
FT FT 189753..189771  
FT FT /tag= bi  
FT FT /note= "Binds primer 99-5919.pu"  
FT FT 189938..189956  
FT FT /tag= bj  
FT FT /note= "Binds primer 99-5919-215.mis"  
FT FT  
  
Alignment Scores:  
Pred. No.: 1.41e+05 Length: 319608  
Score: 42.00 Matches: 7  
Percent Similarity: 88.89% Conservative: 1  
Best Local Similarity: 77.78% Mismatches: 1  
Query Match: 77.78% Indels: 0  
DB: 22 Gaps: 0  
  
US-10-008-355-25 (1-10) x AAS09301 (1-319608)  
  
QY 2 GlyGlyAsnSerGlySerProValphe 10  
||||||| |:::|||||||  
DB 27135 GGTGGGAATGCTGGGAACCCCTGTGTTT 27109  
  
RESULT 25  
ABQ91212/c  
ID ABQ91212 standard; DNA; 2130 BP.  
XX  
AC ABQ91212;  
XX  
DT 01-OCT-2002 (first entry)  
XX  
DE M. capsulatus gene #1197 for DNA array.  
XX  
KW Micro array; gene; ds; differential expression; gene expression.  
XX  
OS Methylococcus capsulatus.  
XX  
PN W0200255655-A2.  
XX  
PD 18-JUL-2002.  
XX
```

```

PF 14-JAN-2002; 2002WO-N000019.
XX
XX 12-JAN-2001; 2001NO-0000235.
PR 12-JAN-2001; 2001NO-0000239.
XX
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA (TIGR-) TIGR.
XX
XX Birkeland NR, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
XX WPI; 2002-557818/59.
DR
XX
XX Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes -
XX
XX PS Claim 14; Page 497; 678pp; English.
XX
XX The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX invention.
XX
XX SQ Sequence 2130 BP; 363 A; 533 C; 705 G; 429 T; 0 other;

Alignment Scores:
Pred. No.: 852 Length: 2130
Score: 41.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 75.93% Indels: 0
DB: 24 Gaps: 0

US-10-008-355-25 (1-10) x ABQ91212 (1-2130)

QY 1 ThrGlyAsnSerGlySerPro 8
AAK87069
Db 1661 ACCGGCGGCAACGTCGTGCGCG 1638

RESULT 26
AAK87069
XX AAK87069 standard; DNA; 8922 BP.
XX
XX AAK87069;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41881.
DE
XX
XX Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosstatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR

```

```

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR

```



PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis.  
XX  
XX  
PS Disclosure; SEQ ID NO 41881; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and

CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 8922 BP; 2655 A; 1877 C; 2038 G; 2362 T; 0 other;

Alignment Scores:  
Pred. No.: 4.16e+03 Length: 8922  
Score: 41.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 75.93% Indels: 0  
DB: 22 Gaps: 0

US-10-008-355-25 (1-10) x AAK87069 (1-8922)

QY 2 GlyGlyAsnserGlyserProValphe 10

||||| |||||||||

Db 1820 GGAGGTGCCAGTGGAGGCCCTGTCTTT 1846

RESULT 27

ABN25944/C

ID ABN25944 standard; cDNA; 370 BP.

XX AC ABN25944;

XX 24-JUN-2002 (first entry)

XX Human ORFX polynucleotide sequence SEQ ID NO:20365.

Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis; gene; ss.

XX Homo sapiens.

XX WO200192523-A2.

XX PD 06-DEC-2001.

XX PF 29-MAY-2001; 2001WO-US10836.

XX PR 30-MAY-2000; 2000US-206132P.

XX PR 29-AUG-2000; 2000US-228716P.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX P-PSDB; ABP10192.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders  
XX  
XX Disclosure; SEQ ID 20365; 1037pp; English.

CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABLN15762 to ABLN27252 encode the human ORFX  
CC proteins given in ABLP00010 to ABLP1500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 370 BP; 107 A; 66 C; 61 G; 134 T; 2 other;

Alignment Scores:  
Pred. No.: 187 Length: 370  
Score: 40.00 Matches: 7  
Percent Similarity: 70.00% Conservative: 0  
Best Local Similarity: 70.00% Mismatches: 3  
Query Match: 74.07% Indels: 0  
DB: 24 Gaps: 0

US-10-008-355-25 (1-10) x ABLN25944 (1-370)

Oy 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
| | | | | | | | | | | | | | | | | | | | | |  
Db 355 ACCGGGGGAAGAAGGAGAGACCCGCTTTT 326

RESULT 28

ABL78727  
ID ABL78727 standard; cDNA; 450 BP.

XX AC ABL78727;

XX DT 17-MAY-2002 (first entry)

XX DE Human ovarian cancer related cDNA clone SEQ ID NO:1705.

XX KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX OS Homo sapiens.

XX PN WO200192581-A2.

XX PD 06-DEC-2001.

XX PF 29-MAY-2001; 2001WO-US17756.

XX PR 26-MAY-2000; 2000US-207484P.

XX PA (CORI-) CORIXA CORP.

XX PI Algate PA, Harlocker SL, Jones R;

XX WIPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising  
XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
XX polypeptide, antibody specific to polypeptide or T cell expressing

PT polypeptide

XX Claim 1; SEQ ID 1705; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers  
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour  
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to  
CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
CC population of (II), or antigen presenting cells that express (II).  
CC (SI) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
CC sample preferably serum or ovarian tissue. The method comprises  
CC contacting a biological sample from a patient with (IV), detecting the  
CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
CC patient, where the amount of polynucleotide hybridising to (IV) is  
CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
CC specific for an ovarian tumour protein comprising contacting T cells  
CC with (III) or (II). (III) is useful in design and preparation of  
CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
CC and proteins in tumour cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumour cDNA library using well known  
CC techniques.

XX SQ Sequence 450 BP; 148 A; 109 C; 109 G; 81 T; 3 other;

Alignment Scores:  
Pred. No.: 232 Length: 450  
Score: 40.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 74.07% Indels: 0  
DB: 24 Gaps: 0

US-10-008-355-25 (1-10) x ABL78727 (1-450)

Oy 1 ThrGlyGlyAsnSerGlySerPro 8

Db 82 ACAGGAGGTAATTCTGGGAAGCCC 105

RESULT 29

AAD03001

ID AAD03001 standard; DNA; 942 BP.

XX AC AAD03001;

XX DT 31-MAY-2001 (first entry)

XX DE Bacillus licheniformis ACl16 RP-II protease DNA.

XX KW Residual protease II; RP-II; additive; cleaning composition; detergent;  
XX ds.

XX OS Bacillus licheniformis.

XX FH Key Location/Qualifiers

FT CDS 1..942

FT /\*tag= a

FT /product= "Bacillus licheniformis ACl16 RP-II protease"

FT /note= "CDS does not include stop codon"

FT /partial

FT sig\_peptide 1..87

FT /\*tag= b

FT sig\_peptide 88..276

FT /\*tag= c

FT /note= "pro sequence of RP-II protease"

FT mat\_peptide 277..942

FT /\*tag= d

FT /product= "Bacillus licheniformis ACl16 mature RP-II

FT protease"

XX

```

PN WO200116285-A2.
XX
PD 08-MAR-2001.
XX
XX 31-AUG-2000; 2000WO-DK00476.
XX
XX 31-AUG-1999; 99DK-0001212.
PR 20-OCT-1999; 99DK-0001500.
XX
XX (NOVO ) NOVOZYMES AS.
XX
XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;
PI Flensted Lassen S;
XX
XX WPI; 2001-226680/23.
DR P-PSDB; AAE00013.
XX
XX Novel RP-II type protease and its variants useful as constituents in
PT detergent compositions, additives and cleaning compositions -
XX
XX Claim 26; Page 107-108; 132pp: English.
PS
XX The present sequence is Bacillus licheniformis AC116 RP-II (Residual
CC protease II) DNA. RP-II protease is useful as a constituent in additives,
CC detergent compositions and other cleaning compositions, optionally in
CC combination with other enzymes such as proteases, lipases, cellulases,
CC amylases, peroxidases or oxidases. The variants of RP-II have improved
CC properties such as substrate specificities, catalytic rate, stability,
CC especially towards the action of proteolytic enzymes and improved
CC resistance towards peroxidase.
XX
SQ Sequence 942 BP; 259 A; 269 C; 225 G; 189 T; 0 other;

Alignment Scores:
Pred. No.: 525 Length: 942
Score: 40.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 74.07% Indels: 0
DB: 22 Gaps: 0

US-10-008-355-25 (1-10) x AAD03001 (1-942)
Qy 2 GlyGlyAsnSerGlySerProValPhe 10
Db 766 GGAGGTCAAGCGGCTCCCGGTATAT 792

RESULT 30
AAD03003
ID AAD03003 standard; DNA; 954 BP.
XX
XX AAD03003;
XX
XX 31-MAY-2001 (first entry)
DT
DE Bacillus licheniformis CDJ31 RP-II protease DNA.
XX
XX Residual protease II; RP-II; additive; cleaning composition; detergent;
KW ds.
XX
XX Bacillus licheniformis.
OS
XX
XX Key Location/Qualifiers.
FH 1..954
FT CDS /*tag= a
FT /*product= "Bacillus licheniformis CDJ31 RP-II protease"
FT /*note= "CDS does not include stop codon"
FT /*partial
FT sig_peptide 1..84
FT /*tag= b
FT 85..288
FT sig_peptide /*tag= c
FT /*note= "Pro sequence of RP-II protease"
FT

mat_peptide 289..954
/*tag= d
/*product= "Bacillus licheniformis CDJ31 mature RP-II
protease"

WO200116285-A2.
XX
XX 08-MAR-2001.
XX
XX 31-AUG-2000; 2000WO-DK00476.
XX
XX 31-AUG-1999; 99DK-0001212.
PR 20-OCT-1999; 99DK-0001500.
XX
XX (NOVO ) NOVOZYMES AS.
XX
XX Norregaard-madsen M, Rahbek Ostergaard P, Voge Christensen CB;
PI Flensted Lassen S;
XX
XX WPI; 2001-226680/23.
DR P-PSDB; AAE00015.
XX
XX Novel RP-II type protease and its variants useful as constituents in
PT detergent compositions, additives and cleaning compositions -
XX
XX Claim 26; Page 114-115; 132pp: English.
PS
XX The present sequence is Bacillus licheniformis CDJ31 RP-II (Residual
CC protease II) DNA. RP-II protease is useful as a constituent in additives,
CC detergent compositions and other cleaning compositions, optionally in
CC combination with other enzymes such as proteases, lipases, cellulases,
CC amylases, peroxidases or oxidases. The variants of RP-II have improved
CC properties such as substrate specificities, catalytic rate, stability,
CC especially towards the action of proteolytic enzymes and improved
CC resistance towards peroxidase.
XX
SQ Sequence 954 BP; 277 A; 263 C; 215 G; 199 T; 0 other;

Alignment Scores:
Pred. No.: 532 Length: 954
Score: 40.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 74.07% Indels: 0
DB: 22 Gaps: 0

US-10-008-355-25 (1-10) x AAD03003 (1-954)
Qy 2 GlyGlyAsnSerGlySerProValPhe 10
Db 778 GGAGGGCAGAGCGGCTCTCCGTATAT 804

RESULT 31
AAH99612/C
ID AAH99612 standard; cDNA; 1030 BP.
XX
XX AAH99612;
XX
XX 16-OCT-2001 (first entry)
DT
DE Human protein encoding cDNA sequence SEQ ID NO:447.
XX
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnary; antileuc; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;

```



RESULT 33  
AAS19209  
ID AAS19209 standard; cDNA; 1873 BP.  
XX  
XX  
AC AAS19209;  
XX  
XX 09-APR-2002 (first entry)  
XX  
XX DNA encoding human transformer 2-beta protein 29.15.  
XX  
XX Human; transformer 2-beta protein 29.15; gene; cytostatic; haemostatic;  
KW virucide; immunomodulatory; antiinflammatory; malignant tumour;  
KW haemopathy; human immunodeficiency virus; HIV; immunological disease;  
KW inflammation; ss.  
XX  
XX Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 143..940  
FT CDS /\*tag= a  
FT FT /product= "transformer 2-beta protein 29.15"  
XX  
XX WO200192517-A1.  
XX  
XX 06-DEC-2001.  
XX  
XX 21-MAY-2001; 2001WO-CN00838.  
XX  
XX 24-MAY-2000; 2000CN-0115807.  
XX  
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
XX  
XX Mao Y, Xie Y;  
XX  
XX WPI; 2002-083185/11.  
DR P-PSDB; AAU10997.  
XX  
XX Human transformer-2-beta protein 29.15 and encoding polynucleotide,  
PT used in diagnosis and treatment of malignant tumours, haemopathy, human  
PT immunodeficiency virus infection, immunological diseases and  
PT inflammation  
XX  
XX Claim 6; Page 30-31; 36pp; Chinese.  
XX  
XX The invention relates to an isolated polypeptide (I) of human transformer  
CC 2-beta protein 29.15 and the polynucleotide (II) encoding (I). (I) and  
CC (II) are used in diagnosis and treatment of malignant tumour, haemopathy,  
CC human immunodeficiency virus (HIV) infection, immunological diseases and  
CC various inflammations. The present sequence represents the coding  
CC sequence of human transformer 2-beta protein 29.15 as described in the  
CC invention.  
XX  
SQ Sequence 1873 BP; 427 A; 518 C; 541 G; 387 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 1.12e+03 Length: 1873  
Score: 40.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 87.50% Mismatches: 0  
Query Match: 74.07% Indels: 0  
DB: 24 Gaps: 0  
  
US-10-008-355-25 (1-10) x AAS19209 (1-1873)  
  
QY 3 GlyAsnSerGlySerProValPhe 10  
|||||:|||||:|||||:|||||  
Db 1342 GGAACACACCTGGATCACCCTGCTTT 1365  
  
RESULT 34  
AAS29979/c  
ID AAS29979 standard; DNA; 7441 BP.  
XX

AAS29979;  
21-NOV-2001 (first entry)  
Human lung antigen genomic DNA #49.  
Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;  
chicken; sheep; immunosuppressive; antiarthritic; vasotropic;  
antirheumatic; antiproliferative; cytostatic; virucide; neuroprotective;  
cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;  
hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
cerebrovascular disorder; nervous system disorder; bacterial infection;  
fungal infection; viral infection; ocular disorder; endocrine disorder;  
gastrointestinal disorder; renal disorder; respiratory disorder;  
wound healing; skin aging; organ transplantation; food preservative;  
tissue regeneration; anti-infertility; food additive.  
Homo sapiens.  
WO200155303-A2.  
02-AUG-2001.  
17-JAN-2001; 2001WO-US01301.  
31-JAN-2000; 2000US-0179065.  
04-FEB-2000; 2000US-0180628.  
24-FEB-2000; 2000US-0184664.  
02-MAR-2000; 2000US-0186350.  
16-MAR-2000; 2000US-0189874.  
17-MAR-2000; 2000US-0190076.  
18-APR-2000; 2000US-0198123.  
19-MAY-2000; 2000US-0205515.  
07-JUN-2000; 2000US-0209467.  
28-JUN-2000; 2000US-0214886.  
30-JUN-2000; 2000US-0215135.  
07-JUL-2000; 2000US-0216647.  
07-JUL-2000; 2000US-0216880.  
11-JUL-2000; 2000US-0217487.  
11-JUL-2000; 2000US-0217496.  
14-JUL-2000; 2000US-0218290.  
26-JUL-2000; 2000US-0220963.  
26-JUL-2000; 2000US-0220964.  
14-AUG-2000; 2000US-0224518.  
14-AUG-2000; 2000US-0224519.  
14-AUG-2000; 2000US-0225213.  
14-AUG-2000; 2000US-0225214.  
14-AUG-2000; 2000US-0225266.  
14-AUG-2000; 2000US-0225267.  
14-AUG-2000; 2000US-0225268.  
14-AUG-2000; 2000US-0225270.  
14-AUG-2000; 2000US-0225447.  
14-AUG-2000; 2000US-0225757.  
14-AUG-2000; 2000US-0225758.  
14-AUG-2000; 2000US-0225759.  
22-AUG-2000; 2000US-0226279.  
22-AUG-2000; 2000US-0226681.  
22-AUG-2000; 2000US-0226868.  
22-AUG-2000; 2000US-0227182.  
23-AUG-2000; 2000US-0227009.  
30-AUG-2000; 2000US-0228924.  
01-SEP-2000; 2000US-0229287.  
01-SEP-2000; 2000US-0229343.  
01-SEP-2000; 2000US-0229344.  
01-SEP-2000; 2000US-0229345.  
05-SEP-2000; 2000US-0229509.  
05-SEP-2000; 2000US-0229513.  
06-SEP-2000; 2000US-0230437.  
06-SEP-2000; 2000US-0230438.  
08-SEP-2000; 2000US-0231242.  
08-SEP-2000; 2000US-0231243.  
08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-457723/49.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT respiratory disorders related to the lung including lung cancers and  
PT also for testing and detection e.g. diagnosis -  
XX  
XX Claim 1; SEQ ID NO 243; 507pp; English.  
XX  
XX Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode  
CC the lung antigen polypeptides of the invention. Lung antigen polypeptides  
CC and their associated polynucleotides are useful in the diagnosis,  
CC treatment and prevention of various types of disorders in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A  
CC pathological condition can be determined by detecting the presence or  
CC absence of a mutation in a lung antigen polynucleotide. The treatable  
CC disorders include autoimmune diseases such as rheumatoid arthritis,  
CC hyperproliferative disorders such as neoplasms of the breast or liver,  
CC cardiovascular disorders such as cardiac arrest, cerebrovascular  
CC disorders such as cerebral ischaemia, nervous system disorders such as  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
CC ocular disorders such as corneal infection, endocrine disorders such as  
CC premature labour and infertility, gastrointestinal disorders such as  
CC Crohn's disease, renal disorders such as glomerulonephritis and  
CC respiratory disorders such as asthma and pleurisy. The polypeptides can  
CC also be used to aid wound healing, to prevent skin aging due to sunburn,  
CC to maintain organs before transplantation, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Alignment Scores:  
Pred. No.: 5.17e+03 Length: 7441  
Score: 40.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 87.50% Mismatches: 0  
Query Match: 74.07% Indels: 0  
DB: 22 Gaps: 0

US-10-008-355-25 (1-10) x AAS29979 (1-7441)  
QY 2 GlycylAsnSerGlySerProVal 9  
|||||:|||||  
Db 2023 GCGGTAAACACAGCAGCCCTGTG 2000  
RESULT 35  
AAS20000  
ID AAS20000 standard; DNA: 44861 BP.  
XX  
AC AAS20000;  
XX







PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
PT  
XX  
XX Claim 1; SEQ ID NO 9954; 530pp; English.  
PS  
XX The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging, the  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 584 BP; 154 A; 173 C; 123 G; 134 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 470 Length: 584  
Score: 39.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 75.00% Mismatches: 0  
Query Match: 72.22% Indels: 0  
DB: 22 Gaps: 0  
US-10-008-355-25 (1-10) x ABA31488 (1-584)  
QY 3 GlyAsnSerGlySerProValPhe 10  
DB 160 GGGAACTCTGGGACCCCATATTT 137  
RESULT 38  
AAK12811/c  
ID AAK12811 standard; DNA; 584 BP.  
XX  
XX AAK12811;  
AC  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 12802.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157275-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR  
XX 26-MAY-2000; 2000US-0207456.  
PR  
XX 30-JUN-2000; 2000US-0608408.  
PR  
XX 03-AUG-2000; 2000US-0632366.  
PR  
XX 21-SEP-2000; 2000US-0234687.  
PR  
XX 27-SEP-2000; 2000US-0236359.  
PR  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
PT  
XX  
XX Example 4; SEQ ID NO: 12802; 650pp + Sequence Listing; English.  
PS  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
XX SQ Sequence 584 BP; 154 A; 173 C; 123 G; 134 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 470 Length: 584  
Score: 39.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 75.00% Mismatches: 0  
Query Match: 72.22% Indels: 0  
DB: 22 Gaps: 0  
US-10-008-355-25 (1-10) x AAK12811 (1-584)  
QY 3 GlyAsnSerGlySerProValPhe 10  
DB 160 GGGAACTCTGGGACCCCATATTT 137  
RESULT 39  
AAK1488/c  
ID AAK1488 standard; cDNA; 741 BP.  
XX  
XX AAK14488;  
AC  
XX  
XX 13-MAR-2001 (first entry)  
DT  
XX  
DE Aspergillus oryzae EST SEQ ID NO:7011.  
XX  
XX Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
XX Aspergillus oryzae.  
OS  
XX  
XX WO2000056762-A2.  
PN  
XX  
XX 28-SEP-2000.  
PD  
XX  
XX 22-MAR-2000; 2000WO-US07781.  
PF  
XX  
XX 22-MAR-1999; 99US-0273623.  
PR  
XX  
XX (NOVO ) NOVO NORDISK BIOTECH INC.  
PA (NOVO ) NOVO NORDISK AS.  
XX  
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
PI WPI; 2000-594572/56.  
DR  
XX  
XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags -  
XX  
XX Claim 88; Page 2846-2847; 3161pp; English.  
XX

CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from  
 CC *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus*  
 CC *AAFL4879* to AAF15337 represents ESTs from *Trichoderma reesei*, which are  
 CC all specifically claimed in the present invention.

XX SQ Sequence 741 BP; 162 A; 244 C; 193 G; 142 T; 0 other;

Alignment Scores:  
 Pred. No.: 611 Length: 741  
 Score: 39.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 72.22% Indels: 0  
 DB: 21 Gaps: 0

US-10-008-355-25 (1-10) x AAF14488 (1-741)

QY 1 ThrGlyGlyAsnSerGlySerProVal 9

DB 339 ACCGGTGGCGTAAGCGGAGCCAGTT 313

RESULT 40

ABK32891/C

ID ABK32891 standard; DNA; 1173 BP.

XX AC ABK32891;

DT 23-APR-2002 (first entry)

DE DNA encoding C. albicans CNS1 target for antifungal compound.

XX antifungal; fungal gene transcription; RPC34; POP3; TFA2; NAB2;

KW MPT1; MTR2; BOS1; POL30; RSA2; SQT1; MTW1; TFB1; SPC98; BFR2; RNA1;

KW GCD7; SK16; NIP1; LCP5; NCE103; ECOL; ORC2; CNS1; YPD1; TIM10; SRB4;

KW yeast; fungus; ds; gene.

XX Candida albicans.

OS WO200202055-A2.

XX 10-JAN-2002.

PF 28-JUN-2001; 2001WO-US20592.

PR 29-JUN-2000; 2000US-215164P.

PR 10-AUG-2000; 2000US-224457P.

PA (ANAD-) ANADYS PHARM INC.

XX Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;

PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long F;

PI Davidov E, Thompson CM;

XX

DR WPI; 2002-147962/19.

XX P-PSDB; AAU83003.

XX Screening candidate antifungal compound for interaction with essential  
 PT protein, modulation of essential protein activity, binding to essential  
 PT protein, by contacting protein with test compound and determining  
 XX effects

PS Disclosure; Figure 80; 522pp; English.

XX The invention describes a method of screening a candidate antifungal  
 CC compound for interaction with essential proteins (EP) or for modulation  
 CC of EP activity e.g fungal gene transcription. The proteins tested in the  
 CC invention include RPC34, POP3, TFA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,  
 CC SQT1, MTW1, TFB1, SPC98, BFR2, RNA1, GCD7, SK16, NIP1, LCP5, NCE103,  
 CC ECOL, ORC2, CNS1, YPD1, TIM10 and SRB4 from *S. cerevisiae*, *C. albicans*  
 CC and human homologues. The method involves contacting a culture with one  
 CC or more test compounds and determining the effects on the growth or  
 CC viability of the culture of cells which preferably comprises fungal cells  
 CC or yeast cells. Preferably the identified compounds interact with, or  
 CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor  
 CC compounds identified by the method are useful for preventing or  
 CC inhibiting fungal, particularly C. albicans growth in culture or in a  
 CC mammal. The antifungal agents interact with essential fungal elements  
 CC that can be used to treat fungal infection by preventing the growth and  
 CC preferentially killing the fungi, but does not inhibit the biological  
 CC activity of mammalian homologues. This sequence encodes a target protein  
 CC used to test the antifungal compounds, described in the method of the  
 CC invention.

XX SQ Sequence 1173 BP; 445 A; 180 C; 239 G; 309 T; 0 other;

Alignment Scores:

Pred. No.: 1.02e+03 Length: 1173  
 Score: 39.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 72.22% Indels: 0  
 DB: 24 Gaps: 0

US-10-008-355-25 (1-10) x ABK32891 (1-1173)

QY 2 GlyGlyAsnSerGlySerPro 8

DB 101 GGAGGTAATTCAGGTCACCT 81

RESULT 41

AAK04382

ID AAK04382 standard; DNA; 1261 BP.

XX AC AAK04382;

DT 13-APR-1999 (first entry)

DE Human secreted protein gene 72 clone HCFNN01.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

XX WO9856804-A1.

PN 17-DEC-1998.

PD 11-JUN-1998; 98WO-US12125.

XX



KW Human; breast specific gene; breast cancer; gene therapy; breast disease;  
KW cytostatic; gene; ds.  
XX Homo sapiens.  
OS  
XX WO200240672-A2.  
PN  
XX  
XX  
XX 23-MAY-2002.  
PD  
XX  
XX 20-NOV-2001; 2001WO-US45079.  
PF  
XX  
XX 20-NOV-2000; 2000US-249998P.  
PR  
XX  
XX 22-NOV-2000; 2000US-252563P.  
PR  
XX  
XX (DIAD-) DIADEXUS INC.  
PA  
XX Salceda S, Macina RA, Recipon H, Caferkey R, Sun Y, Liu C;  
XX  
XX WPI: 2002-500220/53.  
XX  
XX Novel breast-specific polypeptides and polynucleotides encoding  
PT polypeptide, useful for identifying, diagnosing, monitoring, staging,  
PT imaging and treating breast cancer and non-cancerous disease states in  
PT breast  
XX  
XX Claim 1; Page 207-208; 243pp; English.  
PS  
XX The present invention provides human breast specific coding sequences and  
CC proteins. These are useful for detecting breast tissue and for detecting  
CC and treating breast cancer and other breast diseases. The present  
CC sequence is a breast specific coding sequence of the invention.  
XX  
XX Sequence 1775 BP; 570 A; 353 C; 256 G; 596 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 1.61e+03 Length: 1775  
Score: 39.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 72.22% Indels: 0  
DB: 24 Gaps: 0  
US-10-008-355-25 (1-10) x ABT03070 (1-1775)  
QY 2 GlyGlyAsnSerGlySerPro 8  
Db 396 GGGGGGAATTCAGGCTCACCA 416  
RESULT 44  
AAC59942  
-ID AAC59942 standard; cDNA; 2012 BP.  
XX AAC59942;  
AC  
XX 30-JAN-2001 (first entry)  
XX  
XX Human secreted protein cDNA sequence #36.  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200055198-A1.  
PN  
XX  
XX 21-SEP-2000.  
PD  
XX  
XX 09-MAR-2000; 2000WO-US06012.  
PF  
XX  
XX 12-MAR-1999; 99US-0124093.  
PR

PR 23-NOV-1999; 99US-0166989.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Rosen CA, Ruben SM, Komatsoulis G;  
PI  
XX WPI: 2000-587520/55.  
DR  
XX P-PSDB; AAB28737.  
DR  
XX  
XX Human secreted proteins and the nucleic acids that encode them, useful  
PT in gene therapy protocols and recombinant nucleic acid based procedures  
PT  
XX  
XX Claim 1; Page 336; 391pp; English.  
PS  
XX  
XX The invention relate to the isolation of genes AAC59907-C59956 encoding  
CC 50 human secreted proteins AAB28702-B28751. The genes can be used to  
CC generate fusion proteins by linking to the gene for the human  
CC immunoglobulin G Fc portion for increasing the stability of  
CC the fusion protein as compared to the human protein only. The genes and  
CC proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated  
CC from a range of human tissues disclosed in the specification. The  
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
XX  
XX Sequence 2012 BP; 629 A; 406 C; 330 G; 647 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 1.85e+03 Length: 2012  
Score: 39.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 72.22% Indels: 0  
DB: 21 Gaps: 0  
US-10-008-355-25 (1-10) x AAC59942 (1-2012)  
QY 2 GlyGlyAsnSerGlySerPro 8  
Db 760 GGGGGGAATTCAGGCTCACCA 780  
RESULT 45  
AAH33087  
-ID AAH33087 standard; cDNA; 2012 BP.  
XX  
XX AAH33087;  
AC  
XX  
XX 03-SEP-2001 (first entry)  
DT  
XX  
XX Human colon cancer antigen encoding cDNA SEQ ID NO:143.  
DE  
XX  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
KW  
XX  
XX Homo sapiens.  
OS  
XX WO200122920-A2.  
PN  
XX  
XX 05-APR-2001.  
PD  
XX  
XX 28-SEP-2000; 2000WO-US26524.  
PF  
XX  
XX 29-SEP-1999; 99US-0157137.  
PR

PR 03-NOV-1999; 99US-0163280.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
XX WPI: 2001-235357/24.  
DR P-PSDB: AAG73656.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
PT  
PS Claim 1; Page 2309; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAH77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 2012 BP; 629 A; 406 C; 330 G; 647 T; 0 other;  
  
Alignment Scores:  
Pred. No.: Length: 2012  
Score: 1.85e+03 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 72.22% Indels: 0  
DB: Gaps: 0  
  
US-10-008-355-25 (1-10) x AAH33087 (1-2012)  
  
QY 2 GlyGlyAsnSerGlySerPro 8  
|||||  
Db 760 GGGGGGATTCAGGCTCACCA 780  
  
RESULT 46  
ABA96872  
ID ABA96872 standard; cDNA: 2063 BP.  
XX  
XX ABA96872;  
AC  
XX  
DT 02-MAY-2002 (first entry)  
DE Human flavoprotein subunit 24-encoding cDNA..  
XX  
XX Human; flavoprotein subunit 24; recombinant production;  
KW malignant tumour; cancer; blood disease; HIV infection; gene therapy;  
KW human immunodeficiency virus; immune disorder; inflammatory condition;  
KW cytostatic; anti-HIV; antiinflammatory; immunomodulator; gene; ss.  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
FH 401..1051  
FT /\*tag= a  
FT /product= "Human flavoprotein subunit 24"  
XX  
PN WO200198488-A1.  
XX

XX 27-DEC-2001.  
PD  
XX  
XX 14-MAY-2001; 2001WO-CN00778.  
PF  
XX  
XX 16-MAY-2000; 2000CN-0115731.  
PR  
XX  
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
PA  
XX  
XX Mao Y, Xie Y;  
PI  
XX WPI: 2002-090438/12.  
XX P-PSDB: AAM49089.  
DR  
XX Human flavoprotein subunit 24 and encoding polynucleotide, used in  
PT diagnosis and treatment of malignant tumors, hemopathy, human  
PT immunodeficiency virus infection, immunological diseases and  
PT inflammation -  
XX  
PS Claim 6; Page 29-30; 35pp; Chinese.  
XX  
XX This sequence represents cDNA encoding human flavoprotein subunit 24. The  
CC protein has a molecular weight of 24 kD. The invention relates to human  
CC flavoprotein subunit 24 (AAM49089), nucleic acids encoding it (ABA96872),  
CC and a method for the recombinant production of flavoprotein subunit 24.  
CC The present invention additionally discloses an antagonist of  
CC flavoprotein subunit 24 for therapeutic use, and an antibody which  
CC specifically binds to flavoprotein subunit 24. Flavoprotein subunit 24,  
CC and nucleotides which encode it may be used for treating a variety of  
CC diseases, such as malignant tumours, blood diseases, HIV (human  
CC immunodeficiency virus) infection, immune disorders and inflammatory  
CC conditions. The protein may also be used to screen for modulators of its  
CC activity or for peptide fingerprinting identification. The polynucleotide  
CC can be used as a primer for nucleic acid amplification reactions or as a  
CC probe for hybridisation reactions, or in producing gene chips or  
CC microarrays.  
XX  
SQ Sequence 2063 BP; 682 A; 372 C; 447 G; 562 T; 0 other;  
  
Alignment Scores:  
Pred. No.: Length: 2063  
Score: 1.9e+03 Matches: 7  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 87.50% Mismatches: 0  
Query Match: 72.22% Indels: 0  
DB: Gaps: 0  
  
US-10-008-355-25 (1-10) x ABA96872 (1-2063)  
  
QY 1 ThrGlyGlyAsnSerGlySerPro 8  
|||||  
Db 843 ACAGGAGGATCTCTGGATCACCC 866  
  
RESULT 47  
ABL90832  
ID ABL90832 standard; cDNA: 2107 BP.  
XX  
XX ABL90832;  
AC  
XX  
XX 24-MAY-2002 (first entry)  
DT  
XX  
XX Human polynucleotide SEQ ID NO 1394.  
DE  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiathergenic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200190304-A2.  
XX

PD 29-NOV-2001.  
XX 18-MAY-2001; 2001WO-US16450.  
XX 19-MAY-2000; 2000US-205515P.  
PR (HUMA-) HUMAN GENOME SCI INC.  
XX Birse CE, Rosen CA;  
XX WPI: 2002-122018/16.  
XX P-PSDB; ABB90423.  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
XX prevention of neural, immune system, muscular, reproductive,  
XX gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
XX disorders -  
XX Claim 4; SEQ ID NO 1394; 2081pp + Sequence Listing; English.  
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
XX (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful  
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX and parasitic infections.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 2107 BP; 488 A; 514 C; 558 G; 545 T; 2 other;  
  
Alignment Scores:  
Pred. No.: 1.94e+03 Length: 2107  
Score: 39.00 Matches: 7  
Percent Similarity: 77.78% Conservative: 0  
Best Local Similarity: 77.78% Mismatches: 2  
Query Match: 72.22% Indels: 0  
DB: 24 Gaps: 0  
  
US-10-008-355-25 (1-10) x ABL90832 (1-2107)  
  
QY 2 GlyGlyAsnSerGlySerProValPhe 10  
||||||| ||| |||||  
Db 1807 GGGGGACGCTGGGGACCTGTTT 1833  
  
RESULT 48  
AAS27076  
ID AAS27076 standard; cDNA: 2109 BP.  
XX AAS27076;  
AC AAS27076;  
XX 07-NOV-2001 (first entry)  
XX cDNA encoding novel signal transduction pathway protein, Seq ID 111.  
XX Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;  
KW antineoplastic; anti-HIV; antibacterial; anti-inflammatory; cancer;  
KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
KW organ transplant rejection; infection; hepatitis C; blood disorder;  
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;

KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;  
KW acquired immune deficiency syndrome.  
XX Homo sapiens.  
XX WO200154733-A1.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01312.  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 11-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.  
XX 14-SEP-2000; 2000US-0232401.  
XX 14-SEP-2000; 2000US-0233063.  
XX 14-SEP-2000; 2000US-0233064.  
XX 14-SEP-2000; 2000US-0233065.  
XX 21-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234274.







PA (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-581633/65.  
XX P-PSDB; AAU87408.  
XX New isolated nucleic acid encoding a protein for diagnosing,  
PT preventing, treating or ameliorating medical conditions and used as  
PT food additives or preservatives -  
XX  
XX Claim 1; SEQ ID No 328; 837pp; English.  
PS  
XX The invention describes an isolated nucleic acid molecule (I) encoding a  
CC novel central nervous system protein. (I) and polypeptides (III) encoded  
CC by (I), are used to treat a medical conditions and in diagnosis of a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
CC angioneuroma, nervous system disorders e.g. Alzheimer's disease and  
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
CC adenocarcinomas and irritable bowel syndrome, reproductive system  
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
CC acute kidney failure and blood related disorders e.g. myocardial  
CC infarction. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
Alignment Scores:  
Pred. No.: 1.97e+03 Length: 2136  
Score: 39.00 Matches: 7  
Percent Similarity: 77.78% Conservative: 0  
Best Local Similarity: 77.78% Mismatches: 2  
Query Match: 72.22% Indels: 0  
DB: 23 Gaps: 0  
US-10-008-355-25 (1-10) x ABK43738 (1-2136)  
QY 2 GlyGlyAsnSerGlySerProValphe 10  
Db 1823 GGGGGGAACGTGGGGGAACCTGTGTGT 1849  
RESULT 50  
AAH16791/c  
ID AAH16791 standard; cDNA: 2166 BP.  
XX  
XX AAH16791;  
XX  
XX 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:16029.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
XX  
XX 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI: 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 16029; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination  
CC of the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX SQ Sequence 2166 BP; 646 A; 437 C; 512 G; 571 T; 0 other;  
Alignment Scores:  
Pred. No.: 2e+03 Length: 2166  
Score: 39.00 Matches: 7  
Percent Similarity: 80.00% Conservative: 1  
Best Local Similarity: 70.00% Mismatches: 2  
Query Match: 72.22% Indels: 0  
DB: 22 Gaps: 0  
US-10-008-355-25 (1-10) x AAH16791 (1-2166)  
QY 1 ThrGlyGlyAsnSerGlySerProValphe 10  
Db 1137 ACTGGGGGAACTCCAACTCGCCATCTTT 1108  
Search completed: May 23, 2003, 13:44:09  
Job time : 226 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 11:36:23 ; Search time 1224 Seconds  
(without alignments)  
237.768 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TGENSGSPVF 10

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110500\_3085/app\_query.fasta\_1.199  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=1000  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355.ecgn\_1\_1\_1687\_@runat\_16052003\_110500\_3085 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: \*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*

- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htg\_hum.\*
- 40: em.htg\_mus.\*
- 41: em.htggo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	54	100.0	1289	1	STAGASP
2	54	100.0	1558	6	E03836
3	54	100.0	1586	6	E03835
4	54	100.0	1634	1	SASP
5	54	100.0	3240	1	AF309515
6	54	100.0	5207	1	SWA293885
c 7	54	100.0	290150	1	AF004825
c 8	54	100.0	298050	1	AF003132
c 9	54	100.0	347235	1	AF003361
10	49	90.7	657	1	SEP305145
11	49	90.7	849	6	AX141641
12	49	90.7	3189	1	AF269652
c 13	49	90.7	3189	6	AX144972
c 14	48	88.9	10689	1	AE004008
15	45	83.3	66762	2	AC121125
16	45	83.3	120355	9	AC011338
c 17	45	83.3	175191	9	AC010251
c 18	45	83.3	194355	2	AC116511
c 19	45	83.3	195335	2	AC068065
20	45	83.3	260050	1	AL596166
21	45	83.3	349980	6	AX417039
22	45	83.3	349980	6	AX417042
23	44	81.5	1085	1	EFSPREG
24	44	81.5	1912	9	AF008576
25	44	81.5	2385	10	BC019638
26	44	81.5	36547	2	AC017307
c 27	44	81.5	40221	1	MSGY154
28	44	81.5	58722	2	AC100192
c 29	44	81.5	110000	9	AC119569_2
c 30	44	81.5	162609	9	AC108022
c 31	44	81.5	168683	3	AC009741
c 32	44	81.5	166978	2	AC108085
33	44	81.5	171279	2	AC036185
34	44	81.5	174253	3	AC068590
c 35	44	81.5	178199	3	AC007724
c 36	44	81.5	182056	9	AC073533
37	44	81.5	187036	2	AC124514
38	44	81.5	189014	2	AC051649
39	44	81.5	193946	2	AC121339
40	44	81.5	206924	2	AL845498
41	44	81.5	222941	2	AC102236
c 42	44	81.5	225655	3	AE003695
43	44	81.5	234431	9	AF311103
44	44	81.5	279166	2	AC126675
c 45	44	81.5	346897	1	AP002995
46	43	79.6	534	6	AX431912
47	43	79.6	1448	1	BACGASP
48	43	79.6	1448	6	E03693
49	43	79.6	1448	6	115101
50	43	79.6	62485	9	AL590093

51	43	79.6	70494	2	AC111820	AC111820 Rattus no	124	41	75.9	489	8	AY034773	AY034773 Hordeum b
52	43	79.6	80141	9	HSEMSGAR	Y07848 Homo sapien	125	41	75.9	489	8	AY034774	AY034774 Hordeum b
53	43	79.6	110959	2	AC105728	Rattus no	126	41	75.9	489	8	AY034775	AY034775 Hordeum b
54	43	79.6	122146	9	AC011736	Homo sapi	127	41	75.9	489	8	HB5SDNAA	Z11433 Hordeum bog
55	43	79.6	126312	9	AC000026	Homo sapi	128	41	75.9	489	8	HB031006	U11006 Hordeum bul
56	43	79.6	133240	2	AC120820	Rattus no	129	41	75.9	489	8	HB031009	U11009 Hordeum bul
57	43	79.6	159681	2	AC036239	Homo sapi	130	41	75.9	489	8	HB031010	U11010 Hordeum bul
58	43	79.6	159784	2	AC061996	Homo sapi	131	41	75.9	489	8	HB031015	U11015 Hordeum bul
59	43	79.6	163375	2	AC027777	Homo sapi	132	41	75.9	489	8	HSU031030	U11030 Hordeum spo
60	43	79.6	167960	9	AC093084	Homo sapi	133	41	75.9	489	8	HVU07376	U07376 Hordeum vul
61	43	79.6	168319	9	AC023282	Homo sapi	134	41	75.9	489	8	HVU07391	U07391 Hordeum vul
62	43	79.6	169003	2	AC103169	Rattus no	135	41	75.9	489	8	HVU07394	U07394 Hordeum vul
63	43	79.6	173029	2	AC002059	Homo sapi	136	41	75.9	489	8	HVU07395	U07395 Hordeum vul
64	43	79.6	178273	2	AC005308	Plasmodiu	137	41	75.9	489	8	HVU07396	U07396 Hordeum vul
65	43	79.6	185165	9	AC008732	Homo sapi	138	41	75.9	489	8	HVU07397	U07397 Hordeum vul
66	43	79.6	190957	2	AC123483	Rattus no	139	41	75.9	489	8	HVU07398	U07398 Hordeum vul
67	43	79.6	201572	9	AC007298	Homo sapi	140	41	75.9	489	8	HVU07399	U07399 Hordeum vul
68	43	79.6	226239	2	AC126174	Homo sapi	141	41	75.9	490	8	AF237292	AF237292 Hordeum d
69	43	79.6	329709	1	AP002997	Mesorhizo	142	41	75.9	490	8	AY034689	AY034689 Hordeum c
70	42	77.8	152	6	AX098468	Sequence	143	41	75.9	490	8	AY034690	AY034690 Hordeum c
71	42	77.8	489	8	HB031005	U11005 Hordeum bul	144	41	75.9	490	8	AY034691	AY034691 Hordeum c
72	42	77.8	1443	9	AF096304	Homo sapi	145	41	75.9	490	8	AY034759	AY034759 Hordeum b
73	42	77.8	1542	9	BC012857	Homo sapi	146	41	75.9	490	8	AY034764	AY034764 Hordeum b
74	42	77.8	1580	9	BC009052	Homo sapi	147	41	75.9	491	8	HVU07384	U07384 Hordeum vul
75	42	77.8	2100	9	AF023676	Homo sapi	148	41	75.9	492	8	AY034708	AY034708 Hordeum c
76	42	77.8	5640	1	AB011418	Alteromon	149	41	75.9	503	8	AF027591	AF027591 Hordeum m
77	42	77.8	11136	1	AE000862	Methanoba	150	41	75.9	525	8	AY034776	AY034776 Hordeum b
78	42	77.8	20993	10	AB043785	AB043785 Mus muscu	151	41	75.9	1880	3	AF515834	AF515834 Schistos
79	42	77.8	43320	9	AC005784	Homo sapi	152	41	75.9	1880	3	AF515834	AF515834 Schistos
80	42	77.8	64952	9	AL359745	Human DNA	153	41	75.9	10770	1	AE000025	AE000025 Mycoplasma
81	42	77.8	74073	2	AC111139	Mus muscu	154	41	75.9	10770	1	AE000025	AE000025 Mycoplasma
82	42	77.8	90923	2	AL159157	Homo sapi	155	41	75.9	12858	1	AE005721	AE005721 Caulobact
83	42	77.8	129843	2	AC109760	Rattus no	156	41	75.9	17252	1	MP043738	U43738 Mycoplasma
84	42	77.8	133925	9	AC011444	Homo sapi	157	41	75.9	20341	1	D86418	D86418 Bacillus su
85	42	77.8	145905	9	AC087711	Homo sapi	158	41	75.9	27832	3	CEZK1128	Z47357 Caenorhabdi
86	42	77.8	148228	2	AC051659	Homo sapi	159	41	75.9	47362	9	AC073550	AC073550 Homo sapi
87	42	77.8	167320	2	AF298853	Homo sapi	160	41	75.9	62520	2	AC122564	AC122564 Mus muscu
88	42	77.8	170340	2	AC092703	Homo sapi	161	41	75.9	62927	2	AC091142	AC091142 Homo sapi
89	42	77.8	180892	2	AC097732	Rattus no	162	41	75.9	78260	2	AP005565	AP005565 Oryza sat
90	42	77.8	185334	2	AC129332	Rattus no	163	41	75.9	85341	2	AC115362	AC115362 Rattus no
91	42	77.8	186136	2	AC110433	Rattus no	164	41	75.9	96880	2	AC094310	AC094310 Rattus no
92	42	77.8	186520	2	AC110975	Rattus no	165	41	75.9	125064	2	AC094045	AC094045 Rattus no
93	42	77.8	191645	2	AC124899	Rattus no	166	41	75.9	132668	2	AC129360	AC129360 Rattus no
94	42	77.8	207080	2	AC094453	Rattus no	167	41	75.9	139392	2	AC098829	AC098829 Homo sapi
95	42	77.8	319608	6	AX150825	Sequence	168	41	75.9	150180	2	AC120218	AC120218 Mus muscu
96	42	77.8	323930	1	AP003194	Clostridi	169	41	75.9	150180	2	AC083853	AC083853 Homo sapi
97	42	77.8	342800	1	AP003598	Nostoc sp	170	41	75.9	156515	2	AC037194	AC037194 Homo sapi
98	41	75.9	285	8	AY034698	Hordeum b	171	41	75.9	156596	9	AC090679	AC090679 Homo sapi
99	41	75.9	380	8	AY034760	Hordeum b	172	41	75.9	156786	9	AC095046	AC095046 Homo sapi
100	41	75.9	390	8	HVU07392	Hordeum vul	173	41	75.9	164368	2	AC073986	AC073986 Homo sapi
101	41	75.9	418	8	AY034755	Hordeum b	174	41	75.9	166118	10	AL590629	AL590629 Mouse DNA
102	41	75.9	419	8	HB031031	U11031 Hordeum bul	175	41	75.9	167539	2	AC127761	AC127761 Rattus no
103	41	75.9	443	8	AF237296	Hordeum d	176	41	75.9	167539	2	AC127761	AC127761 Rattus no
104	41	75.9	447	8	AY034765	Hordeum b	177	41	75.9	170883	2	AC102867	AC102867 Mus muscu
105	41	75.9	447	8	AY034766	Hordeum b	178	41	75.9	170924	2	AC087619	AC087619 Homo sapi
106	41	75.9	447	8	AY034767	Hordeum b	179	41	75.9	174599	2	AC106523	AC106523 Rattus no
107	41	75.9	448	8	AY034705	Hordeum c	180	41	75.9	177227	2	AC103580	AC103580 Trypanoso
108	41	75.9	463	11	G22347	human STS	181	41	75.9	179446	2	AC121379	AC121379 Rattus no
109	41	75.9	480	8	AY034667	Hordeum r	182	41	75.9	179446	2	AL845483	AL845483 Mus muscu
110	41	75.9	480	8	AY034680	Hordeum r	183	41	75.9	185066	2	AC095764	AC095764 Rattus no
111	41	75.9	481	8	AY034672	Hordeum r	184	41	75.9	185245	2	AC128046	AC128046 Rattus no
112	41	75.9	481	8	AY034673	Hordeum r	185	41	75.9	186730	2	AC097589	AC097589 Sus scrof
113	41	75.9	481	8	AY034685	Hordeum r	186	41	75.9	188868	2	AC103569	AC103569 Rattus no
114	41	75.9	487	8	AF096723	Hordeum m	187	41	75.9	189662	9	AC015468	AC015468 Homo sapi
115	41	75.9	487	8	AY034693	Hordeum c	188	41	75.9	190801	2	AC114699	AC114699 Rattus no
116	41	75.9	487	8	AY034706	Hordeum c	189	41	75.9	190858	9	AC007537	AC007537 Homo sapi
117	41	75.9	489	8	AF237291	Hordeum d	190	41	75.9	191204	9	AC094926	AC094926 Rattus no
118	41	75.9	489	8	AF237354	Hordeum p	191	41	75.9	192139	9	AC021955	AC021955 Homo sapi
119	41	75.9	489	8	AY034687	Hordeum c	192	41	75.9	192742	2	AC118898	AC118898 Rattus no
120	41	75.9	489	8	AY034696	Hordeum c	193	41	75.9	194143	2	AC022222	AC022222 Homo sapi
121	41	75.9	489	8	AY034739	Hordeum b	194	41	75.9	194916	9	AL138921	AL138921 Human DNA
122	41	75.9	489	8	AY034740	Hordeum b	195	41	75.9	195109	9	AC091906	AC091906 Homo sapi
123	41	75.9	489	8	AY034763	Hordeum b	196	41	75.9	203257	10	AL671971	AL671971 Mouse DNA

197	41	75.9	207571	10	AL670951	AL670951 Mouse DNA	c	270	40	74.1	156392	9	AC026341	AC026341 Homo sapi
198	41	75.9	208430	1	BSUB0005	Z99107 Bacillus su	c	271	40	74.1	156616	2	AC121166	AC121166 Rattus no
199	41	75.9	213190	1	BSUB0004	Z99107 Bacillus su	c	272	40	74.1	158876	2	AC016222	AC016222 Homo sapi
200	41	75.9	215703	2	AL844534	AL844534 Mus muscu	c	273	40	74.1	158892	2	AC108175	AC108175 Bos tauru
201	41	75.9	227208	2	AC096408	AC096408 Rattus no	c	274	40	74.1	159758	9	AP005356	AP005356 Homo sapi
202	41	75.9	235141	9	AC004615	AC004615 Homo sapi	c	275	40	74.1	161235	9	AL139239	AL139239 Human DNA
203	41	75.9	240411	2	AC105428	AC105428 Mus muscu	c	276	40	74.1	161571	9	AC069462	AC069462 Homo sapi
204	41	75.9	293272	2	AC120685	AC120685 Rattus no	c	277	40	74.1	162339	2	AC024151	AC024151 Homo sapi
205	41	75.9	319378	2	AC068780	AC068780 Homo sapi	c	278	40	74.1	164649	9	AC009193	AC009193 Homo sapi
206	40	74.1	400	9	AF308822S2	AF308822 Homo sapi	c	279	40	74.1	165519	3	AC008347	AC008347 Drosophil
207	40	74.1	417	9	HS279075	HS279075 Homo sapi	c	280	40	74.1	165675	9	AC090646	AC090646 Homo sapi
208	40	74.1	595	9	S83549	S83549 Na+/H+ exch	c	281	40	74.1	165802	9	AL135914	AL135914 Human DNA
209	40	74.1	610	11	G55946	G55946 SHGC-101189	c	282	40	74.1	166991	9	AC090884	AC090884 Homo sapi
210	40	74.1	743	6	A86197	A86197 Sequence 85	c	283	40	74.1	167217	9	AC002350	AC002350 Homo sapi
211	40	74.1	743	6	AR155690	AR155690 Sequence	c	284	40	74.1	167237	9	AC007938	AC007938 Homo sapi
212	40	74.1	743	6	E66215	E66215 Genome DNA	c	285	40	74.1	168303	9	AL391683	AL391683 Human DNA
213	40	74.1	864	9	HS330006	HS330006 Homo sapi	c	286	40	74.1	168375	2	AC027525	AC027525 Homo sapi
214	40	74.1	1493	9	HS404888	HS404888 Homo sapi	c	287	40	74.1	169196	9	AC026991	AC026991 Homo sapi
215	40	74.1	1583	9	AF294629	AF294629 Homo sapi	c	288	40	74.1	170018	2	AC109099	AC109099 Rattus no
216	40	74.1	1673	10	MUSTGN38A	D50031 Mouse mrna	c	289	40	74.1	170705	2	AC105328	AC105328 Homo sapi
217	40	74.1	1704	10	BC009143	BC009143 Mus muscu	c	290	40	74.1	170997	2	AC095695	AC095695 Rattus no
218	40	74.1	1770	10	MMAF001465	AF001465 Mus muscu	c	291	40	74.1	171012	9	AC073236	AC073236 Homo sapi
219	40	74.1	1775	9	HS4011376	AJ011376 Homo sapi	c	292	40	74.1	173648	2	AC103002	AC103002 Rattus no
220	40	74.1	2193	1	BSP224479	AJ224479 Bacillus	c	293	40	74.1	175304	2	AC106544	AC106544 Rattus no
221	40	74.1	2265	10	MUSTGN38B	D50032 Mouse mrna	c	294	40	74.1	175986	2	AC023071	AC023071 Homo sapi
222	40	74.1	2711	9	AF073299	AF073299 Homo sapi	c	295	40	74.1	176131	2	AC114887	AC114887 Papio cyn
223	40	74.1	3193	9	AK092346	AK092346 Homo sapi	c	296	40	74.1	177147	9	AC007539	AC007539 Homo sapi
224	40	74.1	5654	9	AB058691	AB058691 Homo sapi	c	297	40	74.1	177263	2	AC040942	AC040942 Homo sapi
225	40	74.1	12509	1	AE008590	AE008590 Rickettsi	c	298	40	74.1	178113	2	AC095116	AC095116 Homo sapi
226	40	74.1	13558	1	AE010007	AE010007 Streptoco	c	299	40	74.1	178498	2	AC095429	AC095429 Rattus no
227	40	74.1	24032	3	CEU28C6	Z68315 Caenorhabdi	c	300	40	74.1	179800	2	AC121450	AC121450 Rattus no
228	40	74.1	40255	3	U41272	U41272 Caenorhabdi	c	301	40	74.1	179866	2	AC096113	AC096113 Rattus no
229	40	74.1	41578	3	CBRG46G14	AC091216 Caenorhab	c	302	40	74.1	179947	2	AC021467	AC021467 Homo sapi
230	40	74.1	62043	9	AL161893	AL161893 Human DNA	c	303	40	74.1	180662	2	AC116044	AC116044 Papio ham
231	40	74.1	77289	2	AC067975	AC067975 Homo sapi	c	304	40	74.1	181231	2	AC016259	AC016259 Homo sapi
232	40	74.1	86519	9	AP001860	AP001860 Homo sapi	c	305	40	74.1	183124	2	AC102274	AC102274 Mus muscu
233	40	74.1	89239	2	AC020126	AC020126 Drosophil	c	306	40	74.1	184831	2	AC016030	AC016030 Homo sapi
234	40	74.1	90380	2	AC096452	AC096452 Rattus no	c	307	40	74.1	185841	2	AC117107	AC117107 Rattus no
235	40	74.1	103259	2	AC012430	AC012430 Homo sapi	c	308	40	74.1	187635	2	AC119381	AC119381 Rattus no
236	40	74.1	104972	9	AC025176	AC025176 Homo sapi	c	309	40	74.1	187681	2	AC080124	AC080124 Homo sapi
237	40	74.1	105570	9	HS49E3	AL160451 Homo sapi	c	310	40	74.1	188283	2	AC113823	AC113823 Rattus no
238	40	74.1	107097	8	AC005727	AC005727 Arabidops	c	311	40	74.1	190737	2	AC019143	AC019143 Homo sapi
239	40	74.1	110000	2	AC003656_3	Continuation (2 of	c	312	40	74.1	192938	2	AC103908	AC103908 Canis fam
240	40	74.1	110000	30	AC078780_1	Continuation (4 of	c	313	40	74.1	195430	2	AC090163	AC090163 Homo sapi
241	40	74.1	111234	9	AC013451	AC013451 Homo sapi	c	314	40	74.1	195470	2	AC079367	AC079367 Mus muscu
242	40	74.1	121077	2	AC105558	AC105558 Rattus no	c	315	40	74.1	196857	10	AL590864	AL590864 Mouse DNA
243	40	74.1	121600	9	HS141H5	AL049176 Human DNA	c	316	40	74.1	197725	2	AL844164	AL844164 Mus muscu
244	40	74.1	126243	2	AC130997	AC130997 Rattus no	c	317	40	74.1	199789	2	AC074168	AC074168 Mus muscu
245	40	74.1	127315	2	AC105109	AC105109 Homo sapi	c	318	40	74.1	200046	2	AL732493	AL732493 Mus muscu
246	40	74.1	128585	2	AC110986	AC110986 Rattus no	c	319	40	74.1	201105	9	AC020703	AC020703 Homo sapi
247	40	74.1	129240	9	AC083826	AC083826 Homo sapi	c	320	40	74.1	201246	10	AL683894	AL683894 Mouse DNA
248	40	74.1	129755	2	AC019296	AC019296 Homo sapi	c	321	40	74.1	210955	10	AC109606	AC109606 Mus muscu
249	40	74.1	129997	10	AL772281	AL772281 Mouse DNA	c	322	40	74.1	211208	2	AC094931	AC094931 Rattus no
250	40	74.1	131824	2	AP004037	AP004037 Oryza sat	c	323	40	74.1	212475	2	AC093021	AC093021 Mus muscu
251	40	74.1	134926	2	AP001347	AP001347 Homo sapi	c	324	40	74.1	212923	2	AC104031	AC104031 Homo sapi
252	40	74.1	138192	2	AC117793	AC117793 Mus muscu	c	325	40	74.1	212940	10	AL591143	AL591143 Mouse DNA
253	40	74.1	140425	9	AC005479	AC005479 Homo sapi	c	326	40	74.1	214557	2	AL844579	AL844579 Mus muscu
254	40	74.1	142126	2	AC125129	AC125129 Mus muscu	c	327	40	74.1	216815	2	AC094986	AC094986 Rattus no
255	40	74.1	144066	2	AC117069	AC117069 Rattus no	c	328	40	74.1	217869	2	AC110176	AC110176 Mus muscu
256	40	74.1	144536	9	AL161454	AL161454 Human DNA	c	329	40	74.1	223879	9	AC008735	AC008735 Homo sapi
257	40	74.1	145540	9	AP001052	AP001052 Homo sapi	c	330	40	74.1	226962	2	AC079635	AC079635 Mus muscu
258	40	74.1	145795	2	AC129835	AC129835 Canis fam	c	331	40	74.1	235375	2	AC125039	AC125039 Mus muscu
259	40	74.1	146549	9	AC087692	AC087692 Homo sapi	c	332	40	74.1	237091	2	AC124760	AC124760 Mus muscu
260	40	74.1	147612	2	AP004883	AP004883 Oryza sat	c	333	40	74.1	237143	9	AC073769	AC073769 Mus muscu
261	40	74.1	147889	2	AC018534	AC018534 Homo sapi	c	334	40	74.1	249839	3	AE003791	AE003791 Drosophil
262	40	74.1	148415	9	AC016670	AC016670 Homo sapi	c	335	40	74.1	251308	2	AC023234	AC023234 Mus muscu
263	40	74.1	148598	9	HSBA51C14	AL121875 Human DNA	c	336	40	74.1	257887	2	AC092193	AC092193 Bos tauru
264	40	74.1	150336	2	AC012670	AL121875 Mus muscu	c	337	40	74.1	273785	1	SME591793	AL591793 Sinorhizo
265	40	74.1	150881	9	AC127830	AC127830 Rattus no	c	338	40	74.1	282610	1	RPXX01	AJ235270 Rickettsi
266	40	74.1	152529	9	AC103854	AC103854 Homo sapi	c	339	40	74.1	303014	2	AC117448	AC117448 Homo sapi
267	40	74.1	152649	9	AC020651	AC020651 Homo sapi	c	340	40	74.1	312830	2	AC055706	AC055706 Homo sapi
268	40	74.1	152831	2	AC113721	AC113721 Rattus no	c	341	40	74.1	340000	9	AP001660	AP001660 Homo sapi
269	40	74.1	156034	9	HSBU16705	AL121938 Human DNA	c	342	40	74.1	340000	9	AP001752	AP001752 Homo sapi

343	39	72.2	231	9	S69373S3	S69375 TAL2-BHLH P	c 416	39	72.2	80540	9	AL807761	AL807761 Human DNA
344	39	72.2	236	5	AF533377	Struthio	417	39	72.2	81268	2	AC016408	AC016408 Homo sapi
345	39	72.2	240	6	AR030257	Sequence	418	39	72.2	83069	2	AC106924	AC106924 Rattus no
346	39	72.2	300	9	HS82C5F	263382 H. sapiens C	c 419	39	72.2	87325	2	AC022948	AC022948 Homo sapi
347	39	72.2	613	9	HS3355A1	613	c 420	39	72.2	87402	9	HSJ437M21	HSJ437M21 Human DNA
348	39	72.2	695	9	HS329056	AJ329056 Homo sapi	421	39	72.2	88695	2	AC017727	AC017727 Drosophil
349	39	72.2	1038	6	AX488876	AX488876 Sequence	c 422	39	72.2	90816	2	AC017493	AC017493 Drosophil
350	39	72.2	1191	14	AF438521	AF438521 Apple ste	c 423	39	72.2	95078	8	AF466931	AF466931 Zea mays
351	39	72.2	1303	6	AX078683	AX078683 Sequence	c 424	39	72.2	98092	2	AC099455	AC099455 Rattus no
352	39	72.2	1414	8	AF156984	AF156984 Colletotr	425	39	72.2	100276	2	AC121222	AC121222 Rattus no
353	39	72.2	1531	8	AGOURAIDD	M90295 Agrocycbe ae	426	39	72.2	100981	2	AC096256	AC096256 Rattus no
354	39	72.2	1567	3	DROEXUA	L22553 Drosophila	c 427	39	72.2	101436	2	AC096978	AC096978 Rattus no
355	39	72.2	1775	6	AX430995	AX430995 Sequence	c 428	39	72.2	102029	2	AC113796	AC113796 Rattus no
356	39	72.2	1793	9	AY007086	AY007086 Homo sapi	c 429	39	72.2	102119	2	AP003919	AP003919 Oryza sat
357	39	72.2	1949	9	IR0050687	AL389984 Homo sapi	c 430	39	72.2	102389	3	AC004546	AC004546 Drosophil
358	39	72.2	2166	9	AK024024	AK024024 Homo sapi	c 431	39	72.2	104146	2	AC117855	AC117855 Rattus no
359	39	72.2	2195	9	AK091636	AK091636 Homo sapi	c 432	39	72.2	108765	2	AF216674	AF216674 Homo sapi
360	39	72.2	2271	5	RCU08605	U08605 Rana catesb	c 433	39	72.2	109185	2	AC127160	AC127160 Rattus no
361	39	72.2	2774	6	AI2248	AI2248 Artificial	434	39	72.2	110000	2	AC114472_1	Continuation (2 of
362	39	72.2	2774	6	II4310	II4310 Sequence 5	435	39	72.2	110000	2	LMFLCHR12_3	Continuation (4 of
363	39	72.2	2774	4	ANPELB	X65552 A.niger pel	436	39	72.2	112600	9	AP005140	AP005140 Homo sapi
364	39	72.2	2934	4	SSU84399	U84399 Sus scrofa	437	39	72.2	114023	2	HS316D5	HS316D5 Homo sapi
365	39	72.2	3176	9	HS801813	AL136845 Homo sapi	c 438	39	72.2	114517	9	AL158152	AL158152 Human DNA
366	39	72.2	3179	9	AK095735	AK095735 Homo sapi	c 439	39	72.2	114791	9	HSJ850E9	HSJ850E9 Human DNA
367	39	72.2	3247	9	AB037828	AB037828 Homo sapi	440	39	72.2	117071	9	AL133375	AL133375 Human DNA
368	39	72.2	3255	6	AX193116	AX193116 Sequence	c 441	39	72.2	127757	9	HSJ1004I9	HSJ1004I9 Human DNA
369	39	72.2	3255	9	AF128536	AF128536 Homo sapi	442	39	72.2	120762	2	AC079551	AC079551 Mus muscu
370	39	72.2	3261	2	AC090234	AC090234 Homo sapi	c 443	39	72.2	121287	9	AC010332	AC010332 Homo sapi
371	39	72.2	3652	8	SCYNL123W	271399 S.cerevisia	c 444	39	72.2	121419	2	AC098554	AC098554 Rattus no
372	39	72.2	3901	9	AK095647	AK095647 Homo sapi	c 445	39	72.2	123526	9	AC005026	AC005026 Homo sapi
373	39	72.2	3917	3	DROGLITAC	L39083 Drosophila	c 446	39	72.2	124767	2	AC108637	AC108637 Rattus no
374	39	72.2	3996	9	AF509494	AF509494 Homo sapi	c 447	39	72.2	124835	2	AP004164	AP004164 Oryza sat
375	39	72.2	5641	14	BMVFL1	X13063 Beet wester	c 448	39	72.2	125097	9	AL691415	AL691415 Human DNA
376	39	72.2	6173	3	SCABLC	L76606 Schistocerc	449	39	72.2	126455	2	AC131451	AC131451 Strongylo
377	39	72.2	8989	6	AX128241	AX128241 Sequence	450	39	72.2	129298	9	AC010638	AC010638 Homo sapi
378	39	72.2	9501	2	AC020848	AC020848 Mus muscu	c 451	39	72.2	129432	9	HS497J21	HS497J21 Human DNA
379	39	72.2	10298	1	AE009596	AE009596 Brucella	c 452	39	72.2	130192	9	AL157832	AL157832 Human DNA
380	39	72.2	10501	1	AE009580	AE009580 Brucella	453	39	72.2	136880	9	AC078787	AC078787 Homo sapi
381	39	72.2	10930	1	SCE126	AL049630 Streptomy	c 454	39	72.2	139550	2	AC110062	AC110062 Homo sapi
382	39	72.2	11274	1	AE005770	AE005770 Caulobact	c 455	39	72.2	140137	2	AC027611	AC027611 Homo sapi
383	39	72.2	11579	1	AE006731	AE006731 Sulfolobu	456	39	72.2	140508	8	OSJN00021	OSJN00021 Human chr
384	39	72.2	12027	1	AE011828	AE011828 Xanthomon	457	39	72.2	141450	9	CNS05T8X	CNS05T8X Oryza sat
385	39	72.2	13302	1	AE001909	AE001909 Deinococc	c 458	39	72.2	141987	2	AC120886	AC120886 Oryza sat
386	39	72.2	13368	1	AE005107	AE005107 Halobact	c 459	39	72.2	141991	2	HSJ821D9	HSJ821D9 Homo sapi
387	39	72.2	16068	9	HS4J4877	AJ224877 Homo sapi	460	39	72.2	142423	2	AC109061	AC109061 Rattus no
388	39	72.2	16975	2	AC012862	AC012862 Drosophil	461	39	72.2	142591	2	AC105854	AC105854 Rattus no
389	39	72.2	18656	9	HS4012824	AT012824 Homo sapi	c 462	39	72.2	143878	2	AL359974	AL359974 Homo sapi
390	39	72.2	20938	2	AC022155	AC022155 Homo sapi	c 463	39	72.2	144153	2	AC125472	AC125472 Oryza sat
391	39	72.2	32608	1	SCBAC36F5	AL592292 Streptomy	c 464	39	72.2	144949	2	AC128930	AC128930 Rattus no
392	39	72.2	37386	8	SPBC1685	AL031154 S.pombe c	c 465	39	72.2	145417	2	AC012303	AC012303 Homo sapi
393	39	72.2	37800	8	SCCXIV38K	269382 S.cerevisia	c 466	39	72.2	146364	9	AC092768	AC092768 Homo sapi
394	39	72.2	41021	8	SPDCJ781L3	AL121994 Human DNA	467	39	72.2	146407	9	AC004953	AC004953 Homo sapi
395	39	72.2	41181	8	SPAC2363	AL138854 S.pombe c	c 468	39	72.2	146857	2	AC092454	AC092454 Homo sapi
396	39	72.2	49261	9	HSB209A2	AL121934 Human DNA	c 469	39	72.2	146881	2	OSJN00287	OSJN00287 Oryza sat
397	39	72.2	55299	3	AC024830	AC024830 Caenorhab	c 470	39	72.2	147224	2	AP003779	AP003779 Homo sapi
398	39	72.2	56765	2	AC131519	AC131519 Rattus no	c 471	39	72.2	147844	2	AC068188	AC068188 Homo sapi
399	39	72.2	56958	2	AC011680	AC011680 Homo sapi	c 472	39	72.2	149132	2	AC011640	AC011640 Homo sapi
400	39	72.2	57963	9	AL139345	AL139345 Human DNA	c 473	39	72.2	149532	2	AC101764	AC101764 Mus muscu
401	39	72.2	59247	2	AC015645	AC015645 Homo sapi	c 474	39	72.2	149791	2	AC107226	AC107226 Oryza sat
402	39	72.2	59268	2	AC130908	AC130908 Rattus no	c 475	39	72.2	149897	9	AC055748	AC055748 Homo sapi
403	39	72.2	64513	2	AC124064	AC124064 Homo sapi	476	39	72.2	150584	2	AC117279	AC117279 Rattus no
404	39	72.2	65526	2	AC100032	AC100032 Mus muscu	c 477	39	72.2	150789	9	HS21F7	HS21F7 Human DNA
405	39	72.2	65684	2	AC130373	AC130373 Homo sapi	478	39	72.2	150972	9	AL512380	AL512380 Human DNA
406	39	72.2	66538	2	AC101615	AC101615 Mus muscu	c 479	39	72.2	151006	9	AC062022	AC062022 Homo sapi
407	39	72.2	67416	2	AC115955	AC115955 Mus muscu	c 480	39	72.2	151847	2	AC022792	AC022792 Homo sapi
408	39	72.2	67416	2	AC115955	AC115955 Mus muscu	c 481	39	72.2	152078	2	AC128821	AC128821 Rattus no
409	39	72.2	67709	2	AC022550	AC022550 Mus muscu	c 482	39	72.2	152355	2	AC120587	AC120587 Rattus no
410	39	72.2	67776	2	AC120383	AC120383 Mus muscu	c 483	39	72.2	153455	2	AC091620	AC091620 Papio cyn
411	39	72.2	70282	2	AC091141	AC091141 Homo sapi	c 484	39	72.2	153855	2	AC022157	AC022157 Homo sapi
412	39	72.2	70282	2	AC091141	AC091141 Homo sapi	c 485	39	72.2	154529	2	AC128212	AC128212 Rattus no
413	39	72.2	70312	2	AC121264	AC121264 Mus muscu	c 486	39	72.2	155269	2	AF191252	AF191252 Homo sapi
414	39	72.2	71521	8	AP000373	AP000373 Arabidops	c 487	39	72.2	155270	9	AL391384	AL391384 Human DNA
415	39	72.2	73656	2	AC027536	AC027536 Homo sapi	488	39	72.2	155366	9	AL104561	AL104561 Homo sapi

489	39	72.2	156119	2	AC073247	Homo sapi	AC073247 Homo sapi	562	39	72.2	179375	3	AC009357	Drosophil
490	39	72.2	156466	2	AL139181	Homo sapi	AL139181 Homo sapi	563	39	72.2	180096	9	AC025483	AC025483 Homo sapi
491	39	72.2	156745	2	AC125846	Rattus no	AC125846 Rattus no	c 564	39	72.2	180306	2	AC024701	AC024701 Homo sapi
492	39	72.2	156775	2	AC120713	Rattus no	AC120713 Rattus no	565	39	72.2	181003	2	AC127549	Mus muscu
493	39	72.2	157242	2	AC083821	Homo sapi	AC083821 Homo sapi	566	39	72.2	181636	3	AC008224	Drosophil
494	39	72.2	157693	2	AC040930	Homo sapi	AC040930 Homo sapi	c 567	39	72.2	181931	2	AC115847	Mus muscu
495	39	72.2	157737	2	AC111959	Rattus no	AC111959 Rattus no	568	39	72.2	182105	2	AC036149	Homo sapi
496	39	72.2	158626	9	AC009033	Homo sapi	AC009033 Homo sapi	569	39	72.2	182222	2	AC011084	Homo sapi
497	39	72.2	158918	2	AL391806	Homo sapi	AL391806 Homo sapi	570	39	72.2	182643	9	CNS01DT7	AL132642 Human chr
498	39	72.2	159231	9	AL161654	Human DNA	AL161654 Human DNA	571	39	72.2	183193	2	AL844554	Mus muscu
499	39	72.2	159295	2	AL713955	Oryza sat	AL713955 Oryza sat	572	39	72.2	185058	2	AC125244	Homo sapi
500	39	72.2	159593	2	AC007924	Homo sapi	AC007924 Homo sapi	573	39	72.2	185331	2	AC125774	Rattus no
501	39	72.2	159729	9	AC018950	Homo sapi	AC018950 Homo sapi	c 574	39	72.2	186135	2	AC124947	Homo sapi
502	39	72.2	160308	2	AP004732	Oryza sat	AP004732 Oryza sat	c 575	39	72.2	186634	2	AC099711	Mus muscu
503	39	72.2	160615	9	AC112236	Homo sapi	AC112236 Homo sapi	576	39	72.2	187023	2	AC006882	Caenorhab
504	39	72.2	160698	2	AC128790	Rattus no	AC128790 Rattus no	c 577	39	72.2	187449	2	AC095750	Rattus no
505	39	72.2	161394	2	AC010759	Homo sapi	AC010759 Homo sapi	c 578	39	72.2	188361	2	AC106686	Rattus no
506	39	72.2	161408	2	AC097910	Rattus no	AC097910 Rattus no	579	39	72.2	189763	9	AC090772	Homo sapi
507	39	72.2	161442	2	AL683875	Homo sapi	AL683875 Homo sapi	c 580	39	72.2	190197	2	AC126172	Homo sapi
508	39	72.2	161742	2	AC036204	Homo sapi	AC036204 Homo sapi	581	39	72.2	190235	2	AC118202	Mus muscu
509	39	72.2	162180	2	AC117782	Mus muscu	AC117782 Mus muscu	582	39	72.2	190640	2	AP001177	Homo sapi
510	39	72.2	162343	2	AC109319	Homo sapi	AC109319 Homo sapi	c 583	39	72.2	190803	9	AL355490	Human DNA
511	39	72.2	163154	2	AC026532	Homo sapi	AC026532 Homo sapi	584	39	72.2	191898	9	AP003071	Homo sapi
512	39	72.2	164184	2	AC079050	Homo sapi	AC079050 Homo sapi	c 585	39	72.2	192001	2	AC068376	Homo sapi
513	39	72.2	164620	9	AC107218	Homo sapi	AC107218 Homo sapi	c 586	39	72.2	192825	2	AC119357	Rattus no
514	39	72.2	164846	2	AC098040	Rattus no	AC098040 Rattus no	c 587	39	72.2	193300	2	AC012534	Homo sapi
515	39	72.2	165652	3	AC092222	Drosophil	AC092222 Drosophil	c 588	39	72.2	193919	2	AC115645	Rattus no
516	39	72.2	165655	9	AL136962	Human DNA	AL136962 Human DNA	c 589	39	72.2	194781	2	AC019124	Homo sapi
517	39	72.2	165971	2	AC120815	Rattus no	AC120815 Rattus no	590	39	72.2	194905	2	AL603764	Homo sapi
518	39	72.2	166509	9	AC069542	Homo sapi	AC069542 Homo sapi	591	39	72.2	196698	2	AC129684	Rattus no
519	39	72.2	166689	2	AC038815	Homo sapi	AC038815 Homo sapi	c 592	39	72.2	196783	2	AL683807	Homo sapi
520	39	72.2	166892	9	AC007463	Homo sapi	AC007463 Homo sapi	593	39	72.2	196944	17	AC106725	Homo sapi
521	39	72.2	166969	2	AC118974	Rattus no	AC118974 Rattus no	c 594	39	72.2	198037	2	AC126523	Rattus no
522	39	72.2	167057	2	AC011188	Homo sapi	AC011188 Homo sapi	595	39	72.2	198217	9	AL354942	Human DNA
523	39	72.2	167133	2	AC016985	Homo sapi	AC016985 Homo sapi	c 596	39	72.2	198591	2	AC117509	Homo sapi
524	39	72.2	167632	2	AC068604	Mus muscu	AC068604 Mus muscu	c 597	39	72.2	198614	2	AC073293	Mus muscu
525	39	72.2	167888	2	AC129805	Homo sapi	AC129805 Homo sapi	598	39	72.2	199002	2	AC090439	Homo sapi
526	39	72.2	168348	2	AC127984	Rattus no	AC127984 Rattus no	c 599	39	72.2	200065	2	AC073939	Mus muscu
527	39	72.2	168503	3	AC115484	Drosophil	AC115484 Drosophil	c 600	39	72.2	200723	9	AC087380	Homo sapi
528	39	72.2	168907	2	AC096097	Rattus no	AC096097 Rattus no	601	39	72.2	201097	2	AC112091	Rattus no
529	39	72.2	169387	2	AC096181	Rattus no	AC096181 Rattus no	602	39	72.2	201673	9	AC021698	Homo sapi
530	39	72.2	169431	2	AC131338	Mus muscu	AC131338 Mus muscu	603	39	72.2	201707	2	AC131309	Homo sapi
531	39	72.2	170078	9	AC015914	Homo sapi	AC015914 Homo sapi	c 604	39	72.2	201728	10	AC011013	Mus muscu
532	39	72.2	170136	9	AL356432	Human DNA	AL356432 Human DNA	605	39	72.2	202804	10	AC021630	Mus muscu
533	39	72.2	170233	2	AC068844	Homo sapi	AC068844 Homo sapi	c 606	39	72.2	203946	2	AC069465	Mus muscu
534	39	72.2	170862	2	AC087784	Homo sapi	AC087784 Homo sapi	c 607	39	72.2	204096	2	AC027704	Homo sapi
535	39	72.2	170943	10	AL596095	Mouse DNA	AL596095 Mouse DNA	608	39	72.2	204644	2	AC122819	Mus muscu
536	39	72.2	171459	2	AC067816	Homo sapi	AC067816 Homo sapi	c 609	39	72.2	205020	2	AC109903	Rattus no
537	39	72.2	171690	9	AC092758	Papio cyn	AC092758 Papio cyn	c 610	39	72.2	205044	2	AC009125	Homo sapi
538	39	72.2	171858	2	AC011672	Homo sapi	AC011672 Homo sapi	611	39	72.2	207651	2	AC127598	Mus muscu
539	39	72.2	172122	2	AC096982	Rattus no	AC096982 Rattus no	612	39	72.2	207877	10	AL645948	Mouse DNA
540	39	72.2	172184	2	AC106105	Rattus no	AC106105 Rattus no	613	39	72.2	209175	10	AL450321	Mouse DNA
541	39	72.2	172352	2	AC103624	Mus muscu	AC103624 Mus muscu	c 614	39	72.2	209729	2	AC116764	Mus muscu
542	39	72.2	174150	2	AC115438	Rattus no	AC115438 Rattus no	615	39	72.2	210467	2	AC095600	Rattus no
543	39	72.2	174380	2	AP003720	Homo sapi	AP003720 Homo sapi	616	39	72.2	210982	10	AL603828	Mouse DNA
544	39	72.2	174445	9	AC051642	Homo sapi	AC051642 Homo sapi	c 617	39	72.2	213281	2	AL683882	Mus muscu
545	39	72.2	175341	2	AC078982	Homo sapi	AC078982 Homo sapi	618	39	72.2	213606	2	AC023811	Mus muscu
546	39	72.2	175406	2	AC087128	Mus muscu	AC087128 Mus muscu	c 619	39	72.2	215995	2	AC116463	Mus muscu
547	39	72.2	175420	2	AC129161	Rattus no	AC129161 Rattus no	c 620	39	72.2	216172	9	AC098973	Homo sapi
548	39	72.2	175456	2	AC118909	Rattus no	AC118909 Rattus no	621	39	72.2	216792	9	AC048352	Homo sapi
549	39	72.2	176096	2	AC111117	Mus muscu	AC111117 Mus muscu	c 622	39	72.2	216954	2	AC128437	Rattus no
550	39	72.2	176161	2	AC123447	Rattus no	AC123447 Rattus no	c 623	39	72.2	218677	2	AC109286	Mus muscu
551	39	72.2	176638	2	AC102137	Rattus no	AC102137 Rattus no	624	39	72.2	219955	9	AF137396	Homo sapi
552	39	72.2	177205	2	AC079217	Mus muscu	AC079217 Mus muscu	625	39	72.2	220877	2	AC024957	Mus muscu
553	39	72.2	177262	2	AC013665	Homo sapi	AC013665 Homo sapi	626	39	72.2	221048	2	AC087329	Mus muscu
554	39	72.2	177748	2	AC016185	Homo sapi	AC016185 Homo sapi	c 627	39	72.2	222186	2	AC103339	Rattus no
555	39	72.2	177850	2	AC095662	Rattus no	AC095662 Rattus no	628	39	72.2	223373	2	AC122881	Mus muscu
556	39	72.2	178053	2	AC021112	Homo sapi	AC021112 Homo sapi	629	39	72.2	224148	2	AC016179	Homo sapi
557	39	72.2	178419	2	AC074040	Mus muscu	AC074040 Mus muscu	c 630	39	72.2	225798	2	AC107014	Homo sapi
558	39	72.2	178419	2	AC079486	Mus muscu	AC079486 Mus muscu	c 631	39	72.2	227543	2	AC094548	Rattus no
559	39	72.2	178783	2	AC079486	Mus muscu	AC079486 Mus muscu	c 632	39	72.2	228421	2	AC124397	Mus muscu
560	39	72.2	178783	2	AC120232	Rattus no	AC120232 Rattus no	c 633	39	72.2	229887	2	AC073297	Mus muscu
561	39	72.2	179052	2	AC120232	Rattus no	AC120232 Rattus no	c 634	39	72.2	230726	2	AL845161	Mus muscu

c 635	39	72.2	231560	2	AC127287	AC127287 Mus muscu	c 708	38	70.4	1766	8	AF325915	AF325915 Euglena 9
c 636	39	72.2	240520	3	AE003648	AE003648 Drosophil	c 709	38	70.4	1784	3	AF115368	AF115368 Ixodes ta
c 637	39	72.2	243029	3	AC130219	AC130219 Mus muscu	c 710	38	70.4	2116	5	PCALN	U03471 Rana catesb
c 638	39	72.2	245499	2	AC016186	AC016186 Homo sapi	c 711	38	70.4	2178	9	AK000077	AK000077 Homo sapi
c 639	39	72.2	249687	2	AC015693	AC015693 Homo sapi	c 712	38	70.4	2266	9	AK095993	AK095993 Homo sapi
c 640	39	72.2	252370	2	AC117214	AC117214 Mus muscu	c 713	38	70.4	2323	9	AK024097	AK024097 Homo sapi
c 641	39	72.2	252984	2	AC063918	AC063918 Homo sapi	c 714	38	70.4	2575	9	AK095491	AK095491 Homo sapi
c 642	39	72.2	266552	2	AC091362	AC091362 Rattus no	c 715	38	70.4	2679	10	BC006021	BC006021 Mus muscu
c 643	39	72.2	293930	2	AC122220	AC122220 Mus muscu	c 716	38	70.4	2685	9	AK001888	AK001888 Homo sapi
c 644	39	72.2	299378	3	AE003588	AE003588 Drosophil	c 717	38	70.4	2838	14	CMVREP	236903 Cocksfoot m
c 645	39	72.2	303626	3	AE003601	AE003601 Drosophil	c 718	38	70.4	2969	8	AF321465	AF321465 Yarrowia
c 646	39	72.2	323635	1	AP003005	AP003005 Mesorhizo	c 719	38	70.4	3005	9	BC016918	BC016918 Homo sapi
c 647	39	72.2	337178	3	PD0SADH08	PD0SADH08 Drosophil	c 720	38	70.4	3083	9	HSMB04520	AL833209 Homo sapi
c 648	39	72.2	348411	1	AP003007	AP003007 Mesorhizo	c 721	38	70.4	3283	6	AR171867	AR171867 Sequence
c 649	39	72.2	349116	1	AP003003	AP003003 Mesorhizo	c 722	38	70.4	3412	6	AR171865	AR171865 Sequence
c 650	38.5	71.3	171890	2	AC097188	AC097188 Rattus no	c 723	38	70.4	3441	9	HSAA14403	AJ414403 Homo sapi
c 651	38.5	71.3	179266	2	AC115261	AC115261 Rattus no	c 724	38	70.4	3482	9	AK056564	AK056564 Homo sapi
c 652	38.5	71.3	337832	2	AC099175	AC099175 Rattus no	c 725	38	70.4	3587	1	D84215	D84215 Vibrio chol
c 653	38	70.4	205	11	KLAJ9960	AJ229960 Kluyverom	c 726	38	70.4	3604	8	AY078951	AY078951 Arabidops
c 654	38	70.4	218	6	AX261764	AX261764 Sequence	c 727	38	70.4	3642	9	BC021290	BC021290 Homo sapi
c 655	38	70.4	222	6	AX261025	AX261025 Sequence	c 728	38	70.4	3667	9	AF057352	AF057352 Homo sapi
c 656	38	70.4	222	6	AX261634	AX261634 Sequence	c 729	38	70.4	4082	14	CFMVSE0	248630 Cocksfoot M
c 657	38	70.4	243	6	AX260963	AX260963 Sequence	c 730	38	70.4	4083	14	AB040447	AB040447 Cocksfoot
c 658	38	70.4	244	11	G43967	G43967 WIAF-3591-S	c 731	38	70.4	4083	14	CKEPRORNA	LA0905 Cocksfoot m
c 659	38	70.4	253	6	AX261078	AX261078 Sequence	c 732	38	70.4	4284	9	AK024507	AK024507 Homo sapi
c 660	38	70.4	273	6	AF295041	AF295041 Fifea aci	c 733	38	70.4	4348	10	MUSSPECTLS	M25236 Mouse cytot
c 661	38	70.4	285	11	G04031	G04031 human SPS W	c 734	38	70.4	4427	6	AX180129	AX180129 Sequence
c 662	38	70.4	351	9	AF174064	AF174064 Homo sapi	c 735	38	70.4	4436	9	HSMB04524	AL833213 Homo sapi
c 663	38	70.4	351	11	G23507	G23507 human SPS W	c 736	38	70.4	4709	9	AK074042	AK074042 Homo sapi
c 664	38	70.4	360	8	AF188055	AF188055 Lembohypl	c 737	38	70.4	5760	1	DVURRF12OR	L16784 Desulfovibr
c 665	38	70.4	360	8	AF188056	AF188056 Camptocha	c 738	38	70.4	7706	9	AE051506	AE051506 Homo sapi
c 666	38	70.4	363	8	AF188053	AF188053 Weymouthia	c 739	38	70.4	7856	1	AE010343	AE010343 Methanopy
c 667	38	70.4	363	8	AF188054	AF188054 Weymouthia	c 740	38	70.4	9689	7	AF151674	AF151674 Bacteriop
c 668	38	70.4	364	8	AF188057	AF188057 Camptocha	c 741	38	70.4	9993	1	AE013408	AE013408 Methanosa
c 669	38	70.4	367	8	AF188051	AF188051 Weymouthia	c 742	38	70.4	10029	1	AE011126	AE011126 Methanosa
c 670	38	70.4	367	8	AF188052	AF188052 Weymouthia	c 743	38	70.4	10095	1	AE009404	AE009404 Agrobacte
c 671	38	70.4	460	1	AF128657	AF128657 Unculture	c 744	38	70.4	10132	1	AE008210	AE008210 Agrobacte
c 672	38	70.4	518	8	AF335680	AF335680 Ballota h	c 745	38	70.4	10837	1	AE010824	AE010824 Methanosa
c 673	38	70.4	531	6	AX302803	AX302803 Sequence	c 746	38	70.4	10965	1	AE002004	AE002004 Delnococc
c 674	38	70.4	538	8	AF335681	AF335681 Marrubium	c 747	38	70.4	11004	1	AE010634	AE010634 Fusobacte
c 675	38	70.4	562	8	AF502027	AF502027 Eremostac	c 748	38	70.4	11279	7	AF151091	AF151091 Propaghe
c 676	38	70.4	563	8	AF502028	AF502028 Gomphoste	c 749	38	70.4	11418	1	AF222789	AF222789 Mycobacte
c 677	38	70.4	563	8	AF502056	AF502056 Stachys o	c 750	38	70.4	11630	2	AC014271	AC014271 Drosophil
c 678	38	70.4	567	8	AF502030	AF502030 Otostegia	c 751	38	70.4	12236	1	AE004409	AE004409 Vibrio ch
c 679	38	70.4	580	8	BTM1CSD7	BTM1CSD7 B. taurus (C	c 752	38	70.4	13347	1	SCC42	AL356992 Streptomy
c 680	38	70.4	590	8	AF327586	AF327586 Prunus av	c 753	38	70.4	13352	1	AE010285	AE010285 Pyrococcu
c 681	38	70.4	598	8	AF115481	AF115481 Prunus sa	c 754	38	70.4	14236	1	AE007107	AE007107 Mycobacte
c 682	38	70.4	598	9	HSAA332725	HSAA332725 Homo sapi	c 755	38	70.4	15675	1	AE007108	AE007108 Mycobacte
c 683	38	70.4	663	6	A00097	A00097 Plasmid PAR	c 756	38	70.4	16652	2	AC014298	AC014298 Drosophil
c 684	38	70.4	663	3	S79777	S79777 [specific D	c 757	38	70.4	19946	9	AL583784	AL583784 Human DNA
c 685	38	70.4	671	5	AF373387	AF373387 Hemidacty	c 758	38	70.4	22688	1	AF312688	AF312688 Mycobacte
c 686	38	70.4	725	4	BTCRABP	X07436 Bovine mRNA	c 759	38	70.4	26821	3	CEY7A9C	299286 Caenorhabdi
c 687	38	70.4	731	9	HSAA339711	AJ339711 Homo sapi	c 760	38	70.4	31934	3	LMFL5883	AL117384 Leishmani
c 688	38	70.4	758	1	AY063623	AY063623 Unculture	c 761	38	70.4	38631	1	MTCY05A6	296072 Mycobacteri
c 689	38	70.4	767	1	AF063712	AF063712 Unculture	c 762	38	70.4	38631	1	MTCY05A6	296072 Mycobacteri
c 690	38	70.4	769	8	AY009807	AY009807 Lembohypl	c 763	38	70.4	39290	2	AC019154	AC019154 Leishmani
c 691	38	70.4	827	8	AF363664	AF363664 Lamium pu	c 764	38	70.4	40221	1	MSGY154	AD000002 Mycobacte
c 692	38	70.4	828	4	BTGPX1	X13684 Bovine gp1	c 765	38	70.4	40965	2	AC012633	AC012633 Homo sapi
c 693	38	70.4	828	6	E40571	E40571 Novel prote	c 766	38	70.4	42206	2	AC130334	AC130334 Homo sapi
c 694	38	70.4	828	6	I09014	I09014 Sequence 3	c 767	38	70.4	50320	2	AC126373	AC126373 Homo sapi
c 695	38	70.4	869	4	OAR18224	OAR18224 Ovis aries	c 768	38	70.4	50320	2	AC126373	AC126373 Homo sapi
c 696	38	70.4	929	4	BTTRYPTMR	X94982 B. taurus mR	c 769	38	70.4	53035	2	AC015843	AC015843 Homo sapi
c 697	38	70.4	930	4	BOVCRABP	M17253 Bovine cell	c 770	38	70.4	53585	2	AC105281	AC105281 Homo sapi
c 698	38	70.4	989	8	AF348561	AF348561 Prunus vi	c 771	38	70.4	60692	2	AC124070	AC124070 Homo sapi
c 699	38	70.4	1002	8	OAR18223	Y18223 Ovis aries	c 772	38	70.4	61226	9	AL139392	AL139392 Human DNA
c 700	38	70.4	1002	8	AF348559	AF348559 Prunus la	c 773	38	70.4	63895	2	AC044880	AC044880 Homo sapi
c 701	38	70.4	1013	8	AF348560	AF348560 Prunus pe	c 774	38	70.4	65321	9	AL359831	AL359831 Human DNA
c 702	38	70.4	1047	4	AB038652	AB038652 Sus scrof	c 775	38	70.4	66110	2	AC113246	AC113246 Mus muscu
c 703	38	70.4	1185	14	AF491929	AF491929 Apple ste	c 776	38	70.4	66539	2	AC095519	AC095519 Rattus no
c 704	38	70.4	1292	9	AF052177	AF052177 Homo sapi	c 777	38	70.4	67137	2	AC117599	AC117599 Mus muscu
c 705	38	70.4	1598	8	ATPYRD	X62909 A. thaliana	c 778	38	70.4	67166	2	AC012546	AC012546 Homo sapi
c 706	38	70.4	1694	10	BC025890	BC025890 Mus muscu	c 779	38	70.4	68179	2	AC116801	AC116801 Homo sapi
c 707	38	70.4	1730	9	BC027922	BC027922 Homo sapi	c 780	38	70.4	69059	2	AC090609	AC090609 Homo sapi



c 781	38	70.4	69207	2	AC124131	Mus muscu	854	38	70.4	121798	9	AF200465	AF200465 Homo sapi
c 782	38	70.4	69226	4	AC123732	Mus muscu	c 855	38	70.4	123160	2	AC129395	AC129395 Rattus no
c 783	38	70.4	69699	2	SSC504726	Mus scrof	856	38	70.4	123239	2	AC102716	AC102716 Mus muscu
c 784	38	70.4	69788	2	AC117741	Mus muscu	857	38	70.4	123300	2	AP000406	AP000406 Homo sapi
c 785	38	70.4	70042	9	AP001424	Homo sapi	c 858	38	70.4	123393	2	AL356912	AL356912 Homo sapi
c 786	38	70.4	70285	2	AC084112	Homo sapi	c 859	38	70.4	125380	2	AL356912	AL356912 Homo sapi
c 787	38	70.4	71927	2	AC130349	Homo sapi	860	38	70.4	125380	2	AC114828	AC114828 Oryza sat
c 788	38	70.4	71948	9	AL831782	Human DNA	c 861	38	70.4	128315	2	AC114828	AC114828 Oryza sat
c 789	38	70.4	73246	2	AC069463	Homo sapi	c 862	38	70.4	128342	9	AL627313	AL627313 Human DNA
c 790	38	70.4	73246	2	AC066592	Homo sapi	c 863	38	70.4	128352	9	AC128104	AC128104 Rattus no
c 791	38	70.4	74759	2	AC100661	Mus muscu	c 864	38	70.4	128654	9	AP001258	AP001258 Homo sapi
c 792	38	70.4	76329	8	AB007648	Arabidops	c 865	38	70.4	128815	2	AL773559	AL773559 Sus scrof
c 793	38	70.4	77841	2	AP000572	Homo sapi	c 866	38	70.4	129758	9	AC118269	AC118269 Homo sapi
c 794	38	70.4	79165	2	AC109979	Rattus no	c 867	38	70.4	129898	9	AC104008	AC104008 Homo sapi
c 795	38	70.4	82289	8	ATAC009895	Arabidops	c 868	38	70.4	130119	2	AC123358	AC123358 Rattus no
c 796	38	70.4	82694	2	AC105819	Rattus no	c 869	38	70.4	130264	2	AC128336	AC128336 Rattus no
c 797	38	70.4	82697	9	AC009363	Homo sapi	c 870	38	70.4	130367	2	AC128945	AC128945 Rattus no
c 798	38	70.4	82988	9	AP000967	Homo sapi	c 871	38	70.4	130753	9	HS51019	HS51019 Homo sapi
c 799	38	70.4	83989	8	H0505A02		c 872	38	70.4	132981	9	HS82J11	HS82J11 Human DNA s
c 800	38	70.4	84113	9	AC003066	Human DNA	c 873	38	70.4	133749	2	AC121622	AC121622 Rattus no
c 801	38	70.4	84114	9	AC003006	Human DNA	c 874	38	70.4	135091	2	AC107062	AC107062 Bos tauru
c 802	38	70.4	85662	9	AC114812	Homo sapi	c 875	38	70.4	136124	9	HS394P21	HS394P21 Homo sapi
c 803	38	70.4	89728	2	AC115316	Rattus no	c 876	38	70.4	137628	9	AC074281	AC074281 Homo sapi
c 804	38	70.4	89939	2	AC022667	Homo sapi	c 877	38	70.4	138536	9	AC116914	AC116914 Homo sapi
c 805	38	70.4	90027	2	AC079522	Mus muscu	c 878	38	70.4	138709	2	AC093395	AC093395 Bos tauru
c 806	38	70.4	91042	9	AC108199	Homo sapi	c 879	38	70.4	138969	9	AC060771	AC060771 Homo sapi
c 807	38	70.4	91071	2	AC022589	Homo sapi	c 880	38	70.4	138982	2	AC104550	AC104550 Mus muscu
c 808	38	70.4	91864	2	AC107442	Rattus no	c 881	38	70.4	139953	9	HSBA504H3	HSBA504H3 Human DNA
c 809	38	70.4	93692	2	AC131064	Homo sapi	c 882	38	70.4	140210	2	AC002993	AC002993 Homo sapi
c 810	38	70.4	95129	2	AC111309	Rattus no	c 883	38	70.4	140385	9	HS20B11	HS20B11 Human DNA
c 811	38	70.4	96006	2	AL139224	Homo sapi	c 884	38	70.4	140582	2	AC128429	AC128429 Rattus no
c 812	38	70.4	96006	9	HSJ590P13		c 885	38	70.4	140597	2	AC128970	AC128970 Rattus no
c 813	38	70.4	97005	2	AC113584	Tetradon	c 886	38	70.4	141701	2	AL671435	AL671435 Homo sapi
c 814	38	70.4	98234	2	AC103302	Rattus no	c 887	38	70.4	142278	2	OSJND0135	OSJND0135 Oryza sat
c 815	38	70.4	99392	8	AC000103	Genomic s	c 888	38	70.4	142623	2	AP005291	AP005291 Oryza sat
c 816	38	70.4	99516	2	AC124874	Rattus no	c 889	38	70.4	142948	2	AL672281	AL672281 Mus muscu
c 817	38	70.4	100000	9	AP000157	Homo sapi	c 890	38	70.4	143039	2	AC125869	AC125869 Rattus no
c 818	38	70.4	101365	2	AP004748	Oryza sat	c 891	38	70.4	143226	9	AL353802	AL353802 Human DNA
c 819	38	70.4	101658	9	AL590005	Human DNA	c 892	38	70.4	143277	2	AC105743	AC105743 Oryza sat
c 820	38	70.4	102739	9	HS671014	Human DNA	c 893	38	70.4	143668	2	AC114336	AC114336 Felis cat
c 821	38	70.4	106729	9	AC005320	Human Chr	c 894	38	70.4	144586	2	AC1354831	AC1354831 Human DNA
c 822	38	70.4	107186	9	AC004757	Homo sapi	c 895	38	70.4	145004	2	AC115416	AC115416 Rattus no
c 823	38	70.4	107311	2	AP003835	Oryza sat	c 896	38	70.4	145911	1	AP003014	AP003014 Mesorhizo
c 824	38	70.4	107374	9	AC115836	Homo sapi	c 897	38	70.4	147274	2	AP005588	AP005588 Oryza sat
c 825	38	70.4	107839	2	AP000574	Homo sapi	c 898	38	70.4	147390	9	AC092645	AC092645 Homo sapi
c 826	38	70.4	109554	2	CNS08C90	Oryza sat	c 899	38	70.4	147952	9	AC008712	AC008712 Homo sapi
c 827	38	70.4	110000	2	AL683889_0	Homo sapi	c 900	38	70.4	147954	2	AP003983	AP003983 Oryza sat
c 828	38	70.4	110000	2	AL683889_1	Continuation (2 of	c 901	38	70.4	148642	9	HS580G13	HS580G13 Human DNA
c 829	38	70.4	110000	2	CEY11388_1	Continuation (2 of	c 902	38	70.4	149286	2	AC021562	AC021562 Homo sapi
c 830	38	70.4	110000	2	LMFLCHR32_11	Continuation (12 of	c 903	38	70.4	149640	2	AC084349	AC084349 Homo sapi
c 831	38	70.4	110000	2	LMFLCHR34_09	Continuation (10 of	c 904	38	70.4	149822	2	AC096526	AC096526 Rattus no
c 832	38	70.4	110000	2	AC020884_1	Continuation (2 of	c 905	38	70.4	149849	9	AC009173	AC009173 Homo sapi
c 833	38	70.4	110000	2	AC024901_1	Continuation (2 of	c 906	38	70.4	149892	2	AC026822	AC026822 Homo sapi
c 834	38	70.4	111321	2	AC026718	Homo sapi	c 907	38	70.4	150326	2	AC128956	AC128956 Rattus no
c 835	38	70.4	112667	9	AC008894	Homo sapi	c 908	38	70.4	150785	2	AC026158	AC026158 Homo sapi
c 836	38	70.4	113807	2	AC102013	Mus muscu	c 909	38	70.4	150835	2	AC128575	AC128575 Rattus no
c 837	38	70.4	113923	2	AC107091	Rattus no	c 910	38	70.4	150962	2	AL627100	AL627100 Homo sapi
c 838	38	70.4	114478	2	AP000489	Homo sapi	c 911	38	70.4	151174	2	AC016189	AC016189 Homo sapi
c 839	38	70.4	115170	9	AC104698	Homo sapi	c 912	38	70.4	151254	2	AP004794	AP004794 Oryza sat
c 840	38	70.4	115205	2	AC119552	Rattus no	c 913	38	70.4	151740	2	AC096136	AC096136 Rattus no
c 841	38	70.4	115519	9	AC105133	Homo sapi	c 914	38	70.4	152263	2	AP005412	AP005412 Oryza sat
c 842	38	70.4	116356	2	AC094938	Rattus no	c 915	38	70.4	152617	2	AC113881	AC113881 Rattus no
c 843	38	70.4	116871	2	AL138738	Homo sapi	c 916	38	70.4	154066	9	HS24018	HS24018 Human DNA
c 844	38	70.4	117072	9	AC010445	Homo sapi	c 917	38	70.4	154080	2	AC024416	AC024416 Homo sapi
c 845	38	70.4	118404	4	AL773591	Pig DNA s	c 918	38	70.4	154243	2	AC055768	AC055768 Homo sapi
c 846	38	70.4	118492	2	AC106566	Rattus no	c 919	38	70.4	154372	2	AC120905	AC120905 Rattus no
c 847	38	70.4	118714	2	AC114243	Rattus no	c 920	38	70.4	155112	2	AL357521	AL357521 Homo sapi
c 848	38	70.4	119389	9	AL450328	Human DNA	c 921	38	70.4	155681	2	AC123431	AC123431 Rattus no
c 849	38	70.4	119760	9	AC025458	Homo sapi	c 922	38	70.4	155958	2	AC091464	AC091464 Mus muscu
c 850	38	70.4	119964	2	AC094470	Rattus no	c 923	38	70.4	155993	2	AC105996	AC105996 Mus muscu
c 851	38	70.4	121127	8	AP004484	Lotus jap	c 924	38	70.4	156206	2	AC105500	AC105500 Rattus no
c 852	38	70.4	121460	2	AC027330	Homo sapi	c 925	38	70.4	156634	8	AC074283	AC074283 Oryza sat
c 853	38	70.4	121584	2	AC008232	Drosophil	c 926	38	70.4	157191	2	AC116681	AC116681 Mus muscu



DEFINITION	Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
DNA encoding V8 protease.	E03836	1	GI:2172050	JP 1992211370-A/2	Staphylococcus aureus	1 (bases 1 to 1558)	Nakamura, E., Tamaki, M., Teraoka, H., Yoshida, N. and Kakudou, S.	NEW PROTEASE	SHIONOGI & CO LTD	
Bacteria; Firmicutes; Bacillales; Staphylococcus.										
REFERENCE										
AUTHORS										
TITLE										
JOURNAL										
COMMENT										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									
PD	03-AUG-1992									
PF	19-FEB-1991	JP 1991024633								
PR	20-FEB-1990	JP 90P 40398								
PI	NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI									MATSUMOTO KOICHI,
SHIONOGI & CO LTD										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									
PD	03-AUG-1992									
PF	19-FEB-1991	JP 1991024633								
PR	20-FEB-1990	JP 90P 40398								
PI	NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI									MATSUMOTO KOICHI,
SHIONOGI & CO LTD										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									
PD	03-AUG-1992									
PF	19-FEB-1991	JP 1991024633								
PR	20-FEB-1990	JP 90P 40398								
PI	NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI									MATSUMOTO KOICHI,
SHIONOGI & CO LTD										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									
PD	03-AUG-1992									
PF	19-FEB-1991	JP 1991024633								
PR	20-FEB-1990	JP 90P 40398								
PI	NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI									MATSUMOTO KOICHI,
SHIONOGI & CO LTD										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									
PD	03-AUG-1992									
PF	19-FEB-1991	JP 1991024633								
PR	20-FEB-1990	JP 90P 40398								
PI	NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI									MATSUMOTO KOICHI,
SHIONOGI & CO LTD										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									
PD	03-AUG-1992									
PF	19-FEB-1991	JP 1991024633								
PR	20-FEB-1990	JP 90P 40398								
PI	NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI									MATSUMOTO KOICHI,
SHIONOGI & CO LTD										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									
PD	03-AUG-1992									
PF	19-FEB-1991	JP 1991024633								
PR	20-FEB-1990	JP 90P 40398								
PI	NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI									MATSUMOTO KOICHI,
SHIONOGI & CO LTD										
OS	Staphylococcus aureus									
PN	JP 1992211370-A/2									

[illegible]

[illegible]

```

DEFINITION   Staphylococcus warneri prom gene, proC gene, proD gene, spw1 gene
              and msrwl gene.
ACCESSION    AJ293885
VERSION      AJ293885.2 GI:22218025
KEYWORDS     Cysteine proteinase; glutamyl endopeptidase; msrwl gene; proC gene;
              proD gene; prom gene; spw1 gene.
SOURCE       Staphylococcus warneri.
ORGANISM     Staphylococcus warneri.
              Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE    1
AUTHORS      Yokoi,K., Kakikawa,M., Kimoto,H., Watanabe,K., Yasukawa,H.,
              Yamakawa,A., Taketo,A. and Kodaira,K.I.
TITLE        Genetic and biochemical characterization of glutamyl endopeptidase
              of Staphylococcus warneri M
JOURNAL      Gene 281 (1-2), 115-122 (2001)
MEDLINE      21623048
REFERENCE    2
AUTHORS      Kakikawa,M.
TITLE        Direct Submission
JOURNAL      Submitted (30-AUG-2000) Kakikawa M., Molecular Biology Group,
              Toyama University, 3190 Gofuku, Toyama, 930-8555, JAPAN
REMARK       revised by author [08-AUG-2002]
COMMENT      On Aug 13, 2002 this sequence version replaced gi:9968802.
FEATURES     Location/Qualifiers
              1..5207
                 /organism="Staphylococcus warneri"
                 /db_xref="taxon:1292"
                 complement(232..708)
                 /gene="msrwl"
                 complement(232..708)
                 /gene="msrwl"
                 /codon_start=1
                 /transl_table=11
                 /protein_id="CAD43735.1"
                 /db_xref="GI:22218026"
                 /translation="MEIVYVAGCLMQVQAFKTIPIGVSTEAGRANGITSNLGEVD
              GYAECKLFDPLSLIRNIMSYLFEVIDYSLNQOQDQIGERYGLISENKHUNE
              AQTFINEREDRAHIVVEVLPSLYVKSAPHEQDRLENYPEDHHLCHIPMHLNRYK"
              991..2148
                 /gene="spw1"
                 991..2148
                 /gene="spw1"
                 /codon_start=1
                 /transl_table=11
                 /protein_id="CAD43736.1"
                 /db_xref="GI:22218027"
                 /translation="MTHHSIKYLIGATITTSALFLTSHEASAAITENNVDGQFNHO
              SNGNATQNOTSTYNDENGPIADVNDNIDNHNQNGDAIVSDNDVDATFNHDE
              AGPLMTANSVDSDFNHQDGNPITANNIDSDFNHDQNGNP IAVDNDNIDNHNQNG
              DAIUSDNDVDAFNDHEDGNPLMTANSVDSDFNHQDGNPIASDNDNIDNHNQNG
              DAIVSDNDTDAFNDHEDGNPLMTANSVDSDFNHQDGNPIASDNDNIDNHNQDQGN
              PIASDNNINNNHNDQNGEAIIVANKNADVGFNHDVGOPIHANNOAMISRSQOQSDHT
              TTPQSAITTHQSNLSPATGQDRTRAPYIGFVITLVAVAFILLRKSARK"
              2432..3382
                 /gene="prom"
                 2432..3382
                 /gene="prom"
                 /codon_start=1
                 /transl_table=11
                 /product="glutamyl endopeptidase"
                 /protein_id="CAC06168.1"
                 /db_xref="GI:9968803"
                 /db_xref="SPTREXBL:Q9FBG1"
                 /translation="MKVKFTASSLLIATLTSATLINPAHAETTSSTDNHOQTQSOQ
              OKTPKIDGNNVPEKERANVILPNNDRHOINDTLGHYAPVTFVQVQNSGTEFTA
              SGVVGCKDLTLNKHVVDAATHGNPRALKAFPSAVNQNPNYGGFTGQITKYGNADL
              AIYKFSNNQNGIGEVVTPATLSDNADTSQNPITVIGYPGDKPLATWESRGKITQ
              IQGDMHYDLSTGGNSGSPVFNSEVIGIHGWGASVNGAVFINNNVQNLKQNI
              EDIHFNSDNDNDNNDNGTDDNNNNNDNNNDNNYDNPDA"
              2620..3378
                 /gene="prom"
                 /product="glutamyl endopeptidase"
                 /evidence=experimental
mat_peptide

```

```

gene
3471..4661
/gene="proC"
3471..4661
/gene="proC"
/codon_start=1
/transl_table=11
/product="cysteine proteinase"
/protein_id="CAD43737.1"
/db_xref="GI:22218028"
/translation="MHHSYKHSIFKLISILLFAILLISTIAFTDLNKSAAQKSNOLIE
INVNDKVPQKONLAQOQYKGYAQLDKQNSQGTGYOLGEAFKIYKFNGEEDNSIYT
XPVLKDGKLIYTLTISPKNEADLKQSKENANYSYKISNFIADLEEIKDKHTNITLT
DEKGFTEENGQVRLVKATPLPTNKNKNSQHNSHLKQOLKATSEPTKINENQAAE
ANDNOVQVENTLKNFKIRQQQDQNSWAGFSAALLNATKNTDKYNAHDMIRLTYPN
VSEELPQFSTNPHQMIKYGNSQGRDIHQSGMPFSYNQIDQLTKDNKGAMILAHSVEN
NPNDPHLGHAMAVVGNKINNEDKLIFWNPWDTLSIQDADSNLLHLSEFRDYTWDDT
MICY"
4698..5024
/gene="proD"
4698..5024
/gene="proD"
/codon_start=1
/transl_table=11
/protein_id="CAD43738.1"
/db_xref="GI:22218029"
/translation="MYOLHFHINDDALTLTKSQDPTIHLFIGNWINPNSAOKSISITP
GVDTNHNOYQILQIDTEHQRIKLTSEDPQLMVLIDYEDTNHIFIQTSVKNSYGTSRP
IRYEKF"
BASE COUNT 1937 a 988 c 751 g 1531 t
ORIGIN

Alignment Scores:
Pred. No.: 45.3 Length: 5207
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x SWA293885 (1-5207)

Qy 1 ThrGlyGlyAsnSerGlySerProvaIphe 10
|||||
Db 3122 ACTGGTGGTAATTCAGGTTCCACCTGATTC 3151

RESULT 7
AP004825/c 290150 bp DNA linear BCT 02-JUL-2002
LOCUS Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,
DEFINITION strain:MW2, section 4/10.
ACCESSION AP004825 BA000033
VERSION AP004825.1 GI:21203989
KEYWORDS Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA.
SOURCE Staphylococcus aureus subsp. aureus MW2
ORGANISM Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1
AUTHORS Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,
Nagai,Y., Iwama,N., Asano,K., Naimi,T., Kuroda,H., Cui,L.,
Yamamoto,K. and Hiramatsu,K.
TITLE Genome and virulence determinants of high virulence
community-acquired MRSA
JOURNAL Lancet 359 (9320), 1819-1827 (2002)
MEDLINE 22040717
PUBMED 12044378
REFERENCE 2 (bases 1 to 290150)
AUTHORS Aoki,K., Oguchi,A., Nagai,Y., Asano,K., Iwama,N., Baba,T.,
Kuroda,M., Hiramatsu,K. and Kikuchi,H.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2002) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/,

```





```
/gene="dltD"
3367. .4542
/gene="dltD"
/note="(glycerophosphate chain)
ORFID:SA0796"
/codon_start=1
/transl_table=11
/product="poly D-alanine transfer protein"
/protein_id="BAB42035.1"
/db_xref="GI:13700739"
/translation="MLKFLPLILISGAVFIVELLPLASPFTGLVNEKTVEDNRTSLT
DQVLKTLQDKLYENKYPYIGSELGKDDPFPAIALNKNANKAPFLGAGGST
DLINAYELASQDKLKKLFTIISQWFTNHGLTNQNFDMARSQTQINQMFOQRNMS
TELKRYAORLLQFPVHNKEYLKYAKNPKEKDSYISGFKENOLIKIEAKSLFAM
DKSPLEHVPATKPDASWDEMOKAVEICKADTSNKGIRDOYWLIIQESKPKVRD
YERNVSPQDLELLVKTMRAGADQYVSPISNGVWYDHGIDKERQAVYKKIHS
TVVNDGGKIYDMTKDYERYIVISDAVHIGWKGVYMDQIAKHKMGEPQPEVDKPRN"
/gene="SA0797"
complement(4805. .5047)
/gene="SA0797"
/gene="SA0797"
/note="ORFID:SA0797"
hypothetical protein, similar to nitrogen fixation protein
NifU"
/codon_start=1
/transl_table=11
/protein_id="BAB42036.1"
/db_xref="GI:13700740"
/translation="MPTEDMFDOQVAEIVTERLPPELLRRDGGCSLIDVEDGIVKLQL
HGACGTCPSSTITLKAGIERALHEEVPGVIEVEQV"
5163. .5471
/gene="SA0798"
5163. .5471
/gene="SA0798"
/note="ORFID:SA0798"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="GI:13700741"
/translation="MVVYGVADVICASCVNAPTSKDIYDWLQPLLLKKYPNISKYTYI
DITKDNLDLHDLQLEFIEREQDELFLPTIMNDEIVADGIYQTKITRIDOKLYNE"
complement(5530. .6594)
/gene="SA0799"
complement(5530. .6594)
/gene="SA0799"
/note="ORFID:SA0799"
hypothetical protein, similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB42038.1"
/db_xref="GI:13700742"
/translation="MRNLVLLGGYGNMRIMSLTSLPQDVTVTLVDRMPHGLKP
EFLYALQKDXDKDRKMPNHPQVNTVYGEINDIDLDQAIVSGNSKIDYDLIITGL
GCDKYHNVPGBAEYTHSIQTLSKARDTHSISELPEGAKGVIGAGLSIELASBLR
ESRSDIYLDYRGPILRNFKPEKSLKYAKFKNVTVPNNTNKVPEGPKYIYND
EPKIDILVYTAGIQPVEVVRNLPIDINSNGRVIVNQYHOVPTYRNVVYVGDCAIPLH
APSAOLAEVGGDIAIVLKKOWLNEPLPKMPKELKVGIVGSLGDKQGFAYIMDRIVT
GRUASLTKSGVWLKYHNG"
6911. .7147
/gene="SA0800"
6911. .7147
/note="ORFID:SA0800"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42039.1"
/db_xref="GI:13700743"
/translation="MNPVIFVCLSNMAKGGDYVFNQLENDPQDVVLEYGCLTHCGICS
AGLVALVNGDIVGSDPEELLQNIYAHKETWIF"
7160. .7519
gene
```

```
/gene="SA0801"
7160. .7519
/gene="SA0801"
/note="ORFID:SA0801"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42040.1"
/db_xref="GI:13700744"
/translation="MPTVILTEAAAYEVKDMLEKANEMPDGYLKIKNVGGCTGLTYCM
GAEAEPEENDEVELEYGLKVLVDKEDAPVLNGTTIDFKQSLMGGGFQIDNPNAISCG
CGSSFTKAVGNAPENC"
7973. .9181
/gene="SA0802"
7973. .9181
/gene="SA0802"
/note="ORFID:SA0802"
hypothetical protein, similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB42041.1"
/db_xref="GI:13700745"
/translation="MAQDRKKVLVLGAGYAGLQTVTKLQKAISTEAEITLINKNEYH
YEATWHEASAGTLNEDVLYPVESLKKDKVNFVQAEVTKIDRDAKKVETNQGIVDF
DILVALGFVSSETFGIEGKDHAFQIENVITARELSRHIEDKFANFAASKEKDDNLS
ILVGGAGFTGVFEFLGELTDRIPELCSKYGVQDNKVKITCTVEAAPKMLPMESEELVNH
```

Alignment Scores:

Pred. No.:	878	Length:	298050
Score:	54.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-10-008-355-25 (1-10) x AP003132 (1-298050)

Qy 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
|||||  
Db 123979 ACTGGTGAACCTCAGGTTCACTGTATTT 123950  
|||||

RESULT 9  
AP003361/c  
LOCUS  
DEFINITION  
Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete  
sequence, section 4/9.  
ACCESSION  
AP003361 BA000017  
VERSION  
AP003361.2 GI:14246761  
KEYWORDS  
SOURCE  
Staphylococcus aureus subsp. aureus Mu50 (sub\_species:aureus Mu50,  
strain:Mu50) DNA.  
ORGANISM  
Staphylococcus aureus subsp. aureus Mu50  
Bacteria; Firmicutes; Bacillales; Staphylococcus.

REFERENCE  
AUTHORS  
1  
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,  
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M.,  
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,  
Mizutani, U. Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C.,  
Sekinizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,  
Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,  
Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.  
Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus  
Lancet 357 (9264), 1225-1240 (2001)  
JOURNAL  
MEDLINE  
21311952  
REFERENCE  
2 (bases 1 to 347235)  
Ohta, T.  
Direct Submission  
Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College  
of Medical Technology and Nursing, Department of Medical  
Technology, 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan  
(E-mail:tohta@sakura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,  
Fax:81-298-53-3454)



COMMENT On May 29, 2001 this sequence version replaced gi:13875305.  
 FEATURES  
 source Location/Qualifiers  
 1. .347235  
 /organism="Staphylococcus aureus subsp. aureus Mu50"  
 /strain="Mu50"  
 /sub\_species="aureus Mu50"  
 /db\_xref="taxon:158878"  
 35. .1021  
 /gene="SAV0992"  
 35. .1021  
 /gene="SAV0992"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oligopeptide ABC transporter ATP-binding protein homologue"  
 /protein\_id="BAB57154.1"  
 /db\_xref="GI:14246762"  
 /translation="MNNVLEVKDLTSLKINNEWLATVENISFELSKGEVLGIVGES  
 GCGKLSLSYIKLLPEKISKLSGSEVIFDGRIDLNEKQQLLDINGNDIAMIFQEPM  
 TALNPVFTIKNQLVESIKSKKSKSEANLAKDLKVGIAKQDEILNSYHQLSGG  
 MRQVMIAWAISSPKLLTADFTALDVTIOAILDLKELQKETQMAIMMTHTDLS  
 VVAEFCDKVLMYAGIVFEGGIKELIHPKHPYTKLLSTIPKKEGOKRLETIEGI  
 VPSIOAFHVNKCFRANRCKKLDICNQSPKMHVCEDVIVRCHLYKNEYKEI"  
 1024. .2004  
 /gene="appf"  
 1024. .2004  
 /genes="appf"  
 /codon\_start=1  
 /transl\_table=11  
 /product="oligopeptide transport system ATP-binding protein AppF homologue"  
 /protein\_id="BAB57155.1"  
 /db\_xref="GI:14246763"  
 /translation="MENILEVNOIKYYIKTKLQKTOYVRAVDVVSFKKGQFQG  
 LVGESCGKSTLGVKVIILEDATSGSIYNGEDITRLQKKLKRKQQQYMFQDPYA  
 SLNPMQVGDISEPLINIKKLPKEIKKEVLLKCVGLSEDAIYKIAHEFSGGQRQ  
 RVGIARALRSLIVADEPVSALDVSQVNLKLDLOEQFKLSYLFIAHLSVVK  
 HTSDVGVMLYGHIVIASDKEIYNPKHPYTKALISSIPOIDKHNRRILIKGELPS  
 PNPPOGCPFHRCPTAKDKCKENIPQLKDIGNHQVACFYVDKVGDLNA"  
 1997. .2959  
 /gene="oppB"  
 1997. .2959  
 /genes="oppB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable oligopeptide transport system permease protein"  
 /protein\_id="BAB57156.1"  
 /db\_xref="GI:14246764"  
 /translation="MLKLILKRLMPLILLISIVVFLAIIQPGDPFSDLONGRIK  
 OEAINAQREKLGNDISPHOYIRWNVHVIHGDGESIKYKRPVIDVIERIPNTILIG  
 AMSLIITYISFALGITSGRYSYSLDTYQVFNLYMLAIPSIAGVFAIFSPFELQ  
 WFFQGSVDLNKEGTFEYMSKIYHTIFPAULGVLISLAGYIOYLRNDLIENSKKDY  
 VLTARSKGLSMNKKYINKHILRNSLPIITFLGADIIVSLGAVITETISYNGIGKLF  
 LESVIOQDYPALMALTLFTSFLLGNLISDITYGIDPIRNS"  
 2971. .3852  
 /gene="SAV0995"  
 2971. .3852  
 /gene="SAV0995"  
 /notes="similar to oligopeptide transport system permease protein OppC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB57157.1"  
 /db\_xref="GI:14246765"  
 /translation="MQNKSQSPFKIALSRFHNKIAMLSIIFLLIITIIVSIIAPLIAP  
 FPNVODDLNKGEMTAONILGTDSCGRONFSRLLYAGRISLSIGTISTGMLLIGIT  
 VGVISYFGSIVDTLLMRITFVYMLFPFLIFALVLAALGDKLNKPYGAILILVAVII  
 VLSWGAIARLVGRKVLQENENEFLLAKSIGTPTYKILIKHLLPNILSVIVQATLLF  
 AGMIVVESGLSFLGFGISKAIPSWGNNLSDAQEGDVIISGKPWIMPPAIMITLTIISI

gene  
 CDS  
 complement(3894. .4883)  
 /gene="trps"  
 complement(3894. .4883)  
 /gene="trps"  
 /note="SAV0996"  
 /codon\_start=1  
 /transl\_table=11  
 /product="tryptophanyl-tRNA synthetase"  
 /protein\_id="BAB57158.1"  
 /db\_xref="GI:14246766"  
 /translation="METLFGIQPSGIPTIIGNYIGALKQFVDVQNDYDCYFCIVDOHA  
 ITPQRLKLRKQTRQAAIYLASGIDPKATLFIQSEVPAHVQAGWMLTTIASGEL  
 EMTQYKDKARQAVEGIPAGLLTYPLMAADIVLYNTINVPVGDDQKQHIETRLNLD  
 RFSRYNDVLYVKPEIRMPKRVGGRVMSLQDPTRMKSKSDNNAKFIISLDEPNVAAKKI  
 KSAVTDSDGIIKFDNRNKPQITNLIYSIYAGLTDMPIKDEAKYEGEGYKFKGDLAEI  
 VRAFLVEFOEKYESFYNSDKLDDILDQGRDKAHKVSFKTVKKMEKAMGLGRK"  
 5177. .5372  
 /gene="SAV0997"  
 5177. .5372  
 /genes="SAV0997"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB57159.1"  
 /db\_xref="GI:14246767"  
 /translation="MVTLFTSPSCTSCRKAKAWLQEHDIPTYTERNIFSEHLTIDEIKQ  
 ILKMTBDGTDEIISTRTKYQKLNVDISLPLQDLYSIQDNPGLLRPIILDNRKLQ  
 VGYNDEIRREFUPRKVYRTFQLQEAQRMVD"  
 5943. .6662  
 /gene="SAV0998"  
 5943. .6662  
 /genes="SAV0998"  
 /notes="similar to negative regulator of genetic competence Meca"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB57160.1"  
 /db\_xref="GI:14246768"  
 /translation="MRIERVDQDTVKLFTYSIEARGFSDRLWNRKRGEFFWWSM  
 MDEINEEDFVVEGFLWIOVHAEEKGEVETISKSKNEDMNMNSDDATQDFQVOEL  
 LAGTLBEGDOLEELFQRTKREKAQSKKSSARKNRTIIVKFNDELDVINYATHS  
 NPITTEFDLLYMVDGTYIYAVHFDVSHVQDEVINDSYQLLEFAYPTDRTREVLYNDYA  
 KIIMSHNVTAQVRRYPPEPTE"  
 6783. .7769  
 /gene="SAV0999"  
 6783. .7769  
 /genes="SAV0999"  
 /notes="similar to transcription factor"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB57161.1"  
 /db\_xref="GI:14246769"  
 /translation="MLVALNEEKERVLTATLRKTOYFCPCGQKQVILKRGLKVISHF  
 AHKLAEOQCFNNETIKYKSKLTLAQMIQQQCKVEIEPFLKEIKQIPDILNNKYV  
 IELOYSPYKQIQLORTGLKMKYKSVLLNDVDYCHNKVKFNHFSILFIPNITRKL  
 HTFNLEKKQIIMFMFOIQYLGCGHKYVAEKRNAKIIIELENAPCDYHAYVKLSKFAINQY  
 IKYCRWQNSVLEPTLSAMVQLQULTDQEVYNYGYIIPQEIYIENHPIEWQLQVLDMLK  
 NKSKLVNDNLNYFKLKKFIVALESKTAIEIKLINNYLNICSDRGNDVQILF"  
 7817. .9625  
 /gene="SAV1000"  
 7817. .9625  
 /gene="SAV1000"  
 /codon\_start=1  
 /transl\_table=11  
 /product="thimet oligopeptidase homologue"  
 /protein\_id="BAB57162.1"  
 /db\_xref="GI:14246770"  
 /translation="MSQQLSREQERKYPEYTWDLTITPKDDEAFEAFAEKEVENEIGK  
 EEOFKRGHIGSAAETLNALEEDTLGKLEKVIYVAHLQKQDQDTTNDKTYGMSRAHQ  
 LIITKFSANSLVPEILQIDEDTIQSFVNSYDKLQKFAFDKLINEKRPHILDAETEK

```

LLTEAQAALSTPSNVGMFSNADVLFEADIDKDGNAHPLTQGTFFIKYLESDDRKLRES
AFRNVKAYGAHNHTLGATLAGVKKNVFNARTHTYKAREKALSNHHPENYVDNLV
KTVHKYLPPLHRYTELKELLGLDDLMKVDLYPLIKDKFEMPYEAEKMLKALEP
MGEYLVNVYKEGLNNKRWVDYENKGRSGGYSGGAHLNTPFILLNWSNTISDIYTLVH
EFGHSAHSFQSFNSDDYTFVAEVASTCNEALLSDYMDKHDDDEKRLLLNOE
LEFRATLFRQTMFAEFHKKIHAIEEAGEPLTPTRMNEEYAKLNKLYFGDSVETDEDI
SKWSRIPHFYNNYVYQATGYSAQSLSHQILTECKPAVDRIYINEFLKKGSSNYPPI
EILKNAGVDMTTPETEQACEVFEQKLNAPKLMKA"
complement(10085...10639)
/gene="yjbM"
CDS
complement(10085...10639)
/gene="yjbM"
/notes="SAV1001"
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase"
/protein_id="BA57163.1"
/db_xref="GI:14246771"
/translation="MKQLSYKVVSKTEKIYTYTCYNEIIPKRDIITESMICDQIONAG
IDLEVPKDDLOKSLKTESKIDLIHAREMEIQAPSLVFFSDVHEEGLKVEGLYPVH
IYTYIINELMGKPIEKNLPPKLETYIQQQOLVTMEELLTIYEWPEKLLNKLKLAIQ
QKTEKLYPDGDFKSMPEKIKSK"
complement(10575...10901)
/gene="SAV1002"

```

## Alignment Scores:

```

Pred. No.: 982 Length: 347235
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

```

US-10-008-355-25 (1-10) x AP003361 (1-347235)

Qy 1 ThrGlyAsnSerGlySerProValPhe 10

|||||

Db 56369 ACTGGTGGTAACCTCAGGTTCACCTGTATTT 56340

## RESULT 10

SEP305145

LOCUS

DEFINITION SEP305145 657 bp DNA linear BCT 25-JAN-2002

Staphylococcus epidermidis partial esp gene for extracellular

serine proteinase.

ACCESSION AJ305145

VERSION AJ305145.1 GI:12580903

KEYWORDS esp gene; extracellular serine proteinase.

SOURCE Staphylococcus epidermidis.

ORGANISM Staphylococcus epidermidis

Bacteria; Firmicutes; Bacillales; Staphylococcus.

## REFERENCE

AUTHORS

TITLE Dubin,G., Chmiel,D., Mak,P., Rakwalaska,M., Rzychon,M. and Dubin,A.

JOURNAL Molecular cloning and biochemical characterisation of proteases

MEDLINE Biol. Chem. 382 (11), 1575-1582 (2001)

REFERENCE 2

PUBMED 11767947

AUTHORS Dubin,G.

TITLE Direct Submission

Submitted (19-JAN-2001) Dubin G., Institute of Molecular Biology,

Jagiellonian University, ul. Mickiewicza 3, Krakow 31-120, POLAND

## FEATURES

source

1. .657

/organism="Staphylococcus epidermidis"

/strain="6746"

/db\_xref="taxon:1282"

1. .654

/gene="esp"

<1. .654

/gene="esp"

/codon\_start=1

/evidence=experimental

```

/transl_table=11
/product="extracellular serine proteinase"
/protein_id="CAC27157.1"
/db_xref="GI:12580904"
/db_xref="SPTREMBL:O9AJX0"
/translation="SVILPNNRHQIFNTTQGHYDAVSFIYIPIDGGYMSGGVVVGGE
NEILTKRHVVNGAKGNPRNISVHPSAKNENDYNGKFGVQEIIPYFGNSDLALIRVSP
NEHNOHIGOVVKVPATISSNTDTRINENITVTGYPGDKPLATMMESYGVKVVYIGGEELR
YDLSTVGGNSGSPVFNKGKQVIGIHYGGVDKNYSVYINDFVQQFLRNNIPDINIQ"
<1. .3
/gene="esp"
sig_peptide
4. .651
mat_peptide
/gene="esp"

```

```

/product="extracellular serine proteinase"
/evidence=experimental

```

BASE COUNT 243 a 88 c 122 g 204 t

## ORIGIN

```

Alignment Scores:
Pred. No.: 59.4 Length: 657
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 1 Gaps: 0

```

US-10-008-355-25 (1-10) x SEP305145 (1-657)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10

|||||

Db 499 GGTGGAAACTCTGGATCTCCAGTATTT 525

## RESULT 11

AX141641

LOCUS

DEFINITION AX141641 849 bp DNA linear PAT 31-MAY-2001

Sequence 363 from Patent WO0134809.

ACCESSION AX141641

VERSION AX141641.1 GI:14281692

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

REFERENCE 1

AUTHORS Kimmerly,W.J.

TITLE Staphylococcus epidermidis nucleic acids and proteins

JOURNAL Patent: WO 0134809-A 363 17-MAY-2001;

GLAXO GROUP LIMITED (GB)

FEATURES Location/Qualifiers

source

1. .849

/organism="synthetic construct"

/db\_xref="taxon:32630"

/notes="synthetic nucleic acid sequence"

BASE COUNT 325 a 122 c 147 g 255 t

## ORIGIN

```

Alignment Scores:
Pred. No.: 71.7 Length: 849
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 6 Gaps: 0

```

US-10-008-355-25 (1-10) x AX141641 (1-849)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10

|||||

Db 694 GGTGGAAACTCTGGATCTCCAGTATTT 720

## RESULT 12

AF269652/c

LOCUS

DEFINITION AF269652 3189 bp DNA linear BCT 01-AUG-2000

Staphylococcus epidermidis strain SRI clone step.1015e12 genomic

```

sequence.
ACCESSION AF269652.1 GI:9623548
VERSION AF269652.1
KEYWORDS Staphylococcus epidermidis
SOURCE Staphylococcus epidermidis
ORGANISM Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1 (bases 1 to 3189)
AUTHORS Kimmerly,W.J., Taylor,J.David., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3189)
AUTHORS Taylor,J.David., Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
Listenbee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
source
1. 3189
/organism="Staphylococcus epidermidis"
/strain="Srl"
/db_xref="taxon:1282"
/cbone="step.1015e12"
BASE COUNT 1081 a 596 c 502 g 1010 t
ORIGIN
Alignment Scores:
Pred. No.: 189 Length: 3189
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 1 Gaps: 0
US-10-008-355-25 (1-10) x AF269652 (1-3189)
Qy 2 GlyGlyAsnSerGlySerProValPhe 10
|||||
Db 1147 GGTGGAACCTCTGGATCCACGATATT 1121
RESULT 13
AX144972/c
LOCUS AX144972 3189 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 3694 from Patent WO0134809.
ACCESSION AX144972
VERSION AX144972.1 GI:14283537
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 3189)
AUTHORS Kimmerly,W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 3694 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
source
1..3189
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"
BASE COUNT 1081 a 596 c 502 g 1010 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 189 Length: 3189
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 6 Gaps: 0
US-10-008-355-25 (1-10) x AX144972 (1-3189)
Qy 2 GlyGlyAsnSerGlySerProValPhe 10
|||||
Db 1147 GGTGGAACCTCTGGATCCACGATATT 1121
RESULT 14
AE004008/c
LOCUS AE004008 10689 bp DNA linear BCT 15-JUN-2001
DEFINITION Xylella fastidiosa 9a5c, section 154 of the complete genome.
ACCESSION AE004008 AE003849
VERSION AE004008.1 GI:9106961
KEYWORDS Xylella fastidiosa 9a5c.
SOURCE Xylella fastidiosa 9a5c.
ORGANISM Bacteria; Proteobacteria; gamma subdivision: Xanthomonas group;
Xylella.
REFERENCE 1 (bases 1 to 10689)
AUTHORS Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvaranga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
Bueno,M.R., Camargo,A., Canargo,L.E., Carraro,D.M., Carret,H.,
Colauto,N.B., Colombo,C., Costa,F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,
Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kempet,E.L., Kitajima,J.P.
and Marino,C.L.
TITLE The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis
JOURNAL Nature 406 (6792), 151-157 (2000)
MEDLINE 20365717
PUBMED 10910347
REFERENCE 2 (bases 1 to 10689)
AUTHORS Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvaranga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A., Camargo,L.E.A., Carraro,D.M.,
Carret,H., Colauto,N.B., Colombo,C., Cristofani,M., Dias-Neto,E.,
Costa-Neto,C.M., Coutinho,L.L., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

```

FEATURES	Location/Qualifiers
source	1. .10689 /organism="Xylella fastidiosa 9a5c" /db_xref="taxon:160492" /clone="9a5c" complement(119. .367) /gene="Xf1871" complement(119. .367) /gene="Xf1871" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84677.1" /db_xref="GI:9106962" METLKLTTESWYPVVGSSLLIAGATAATLFIKLFH" complement(430. .618) /gene="Xf1872" complement(430. .618) /gene="Xf1872" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84678.1" /db_xref="GI:9106963" /translation="MNEEARATAPQGLAASDPQRTDAPKTKGLRKRKPYPTGSASI IVLHVHVTORQQRKASP" 941. .1288 /gene="Xf1873" 941. .1288 /gene="Xf1873" /note="similar to GI:479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF84679.1" /db_xref="GI:9106964" /translation="MSKNKIPNQLTGNWAGQFINGKLVTPERRTIEWQLRWLSLT CTLREWQKMEEARATAPQGLAASDPQRTDALQKTVSHKRRHAANETASIIQLRQ VLTORQQRKASP" 1351. .1662 /gene="Xf1874" 1351. .1662 /gene="Xf1874" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84680.1" /db_xref="GI:9106965" /translation="MTWTLKEDKTEIIDEQKIQIAKLSETAKIQAEIHEVTHQTEM NAHINENNARTQENARTQKLIETQKVSQWYPVVASGLLAAGATAATLFIKLF H" complement(1839. .2918) /gene="Xf1875" complement(1839. .2918) /gene="Xf1875" /note="similar to SP:IP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="phase-related protein" /protein_id="AAF84681.1"
gene	/db_xref="GI:9106966" /translation="MLNLVTGAPGNGKTLAYAVDWLIRQIEIDKSLVKSAGVPRSFVTD TEGFDVEAVRRLTGYYVQSAPEDWETPQGSVIYVDEAHMFPAGRPSDDPRVCDL DTHRHGGYDLMEVTOPTKIHHELRLRAGEHVHLARMCLOTAGLYRMSRAODDPYDI HQREKAEEVWKFPRDRIYALYSTLHVSHKFRIPKPKVMSALSVCVTSVIGLFWHH YYSAPHLSEASAVAGASQASLRAPASLSRSSRLVSGMRTYVLETESAPTLSGCV SSDRSCRCFNTDGYQIDMSVVECRLLASPLPFNVYHAYVTSSSSAAAVSSSPSSPS SSSVLSSSVSAPILGSSSPSSH" complement(2918. .3202) /gene="Xf1876" complement(2918. .3202) /gene="Xf1876" /note="similar to SP:IP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="phase-related protein" /protein_id="AAF84682.1" /db_xref="GI:9106967" /translation="MFSWLAMLLRNLTGOTVARGLVGAGLALVTVPLIPLVTSALNL IVKMSGISADVLNLTALMGFGEALSIIGSAMLTRLALQSLHVGIVKATT" complement(3195. .4412) /gene="Xf1877" complement(3195. .4412) /gene="Xf1877" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84683.1" /db_xref="GI:9106968" /translation="MCSRFFLYFISALGLTVISFPFLFAONVPTVEVVTSVRY SSVISDGVSTATFEARSTALNVRYITVPVDISATLGSLAKAVRRGMATYNTYS MLENGAGWVIDELTHEVMSGPALKEIPVGTVAWMQRPQDCHVFSYTPQGLIAP INQAQSLSPQPLVTSGGSSVSSRWLVHLEGGEYIGHLSTQDSVPDYSGLPP SVILDLQGLVRSDBSTYNAVLTDSQTGAVALTPEIVSALNKLRSLEDELKASHAP DQPPSSGASSPSSSGTAWPFCSNASVVCDDIDWVKDFLKLPLVPDPVVDKL PQAKTSSSGLGEGACPSPTTIPIEFSGYTSVEISYQPPCDFALMRPVVIVATILA AYIAGGFRGVKNV" complement(4481. .4702) /gene="Xf1878" complement(4481. .4702) /gene="Xf1878" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84684.1" /db_xref="GI:9106969" /translation="MKSVIDRGIAFFSLSTSLVFAADAAGASFDGAAVATGALGIS GAVALIGAALAPAAISVGNKWLKGAIFG" complement(4724. .4924) /gene="Xf1879" complement(4724. .4924) /gene="Xf1879" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84685.1" /db_xref="GI:9106970" /translation="MAOQLMSLYCTQYDVEARTCSQQAWMVPSLLPPISYEDVRILL PHIVMCFVLVWGFUFLFTVVRD" complement(4931. .5236) /gene
CDS	656. 48.00
Alignment Scores:	Length: 10689 Pred. No.: 9 Score: Matches:



29174	29273:	gap of	100 bp	
29274	30002:	contig of 729 bp	in length	
30003	30102:	gap of	100 bp	
30103	30825:	contig of 723 bp	in length	
30826	30925:	gap of	100 bp	
30926	31647:	contig of 722 bp	in length	
31648	31747:	gap of	100 bp	
31748	32459:	contig of 712 bp	in length	
32460	32559:	gap of	100 bp	
32560	33277:	contig of 718 bp	in length	
33278	33377:	gap of	100 bp	
33378	34086:	contig of 709 bp	in length	
34087	34186:	gap of	100 bp	
34187	34902:	contig of 716 bp	in length	
34903	35002:	gap of	100 bp	
35003	35714:	contig of 712 bp	in length	
35715	35814:	gap of	100 bp	
35815	36543:	contig of 729 bp	in length	
36544	36643:	gap of	100 bp	
36644	37367:	contig of 724 bp	in length	
37368	37467:	gap of	100 bp	
37468	38191:	contig of 724 bp	in length	
38192	38291:	gap of	100 bp	
38292	39012:	contig of 721 bp	in length	
39013	39112:	gap of	100 bp	
39113	39829:	contig of 717 bp	in length	
39830	39929:	gap of	100 bp	
39930	40639:	contig of 710 bp	in length	
40640	40739:	gap of	100 bp	
40740	41453:	contig of 714 bp	in length	
41454	41533:	gap of	100 bp	
41554	42275:	contig of 722 bp	in length	
42276	42375:	gap of	100 bp	
42376	43090:	contig of 715 bp	in length	
43091	43190:	gap of	100 bp	
43191	43905:	contig of 715 bp	in length	
43906	44005:	gap of	100 bp	
44006	44688:	contig of 683 bp	in length	
44689	44788:	gap of	100 bp	
44789	45493:	contig of 705 bp	in length	
45494	45593:	gap of	100 bp	
45594	46316:	contig of 723 bp	in length	
46317	46416:	gap of	100 bp	
46417	47136:	contig of 720 bp	in length	
47137	47236:	gap of	100 bp	
47237	47927:	contig of 691 bp	in length	
47928	48027:	gap of	100 bp	
48028	48749:	contig of 722 bp	in length	
48750	48849:	gap of	100 bp	
48850	49563:	contig of 714 bp	in length	
49564	49663:	gap of	100 bp	
49664	50375:	contig of 712 bp	in length	
50376	50475:	gap of	100 bp	
50476	51193:	contig of 718 bp	in length	
51194	51293:	gap of	100 bp	
51294	52013:	contig of 720 bp	in length	
52014	52113:	gap of	100 bp	
52114	52887:	contig of 774 bp	in length	
52888	52987:	gap of	100 bp	
52988	53717:	contig of 730 bp	in length	
53718	53817:	gap of	100 bp	
53818	54547:	contig of 730 bp	in length	
54548	54647:	gap of	100 bp	
54648	55374:	contig of 727 bp	in length	
55375	55474:	gap of	100 bp	
55475	56189:	contig of 715 bp	in length	
56190	56289:	gap of	100 bp	

Alignment Scores:  
Pred. No.: 7.34e+03 Length: 66762  
Score: 45.00 Matches: 8  
Percent Similarity: 90.00% Conservative: 1  
Best Local Similarity: 80.00% Mismatches: 1

Query Match:	83.33%	Indels:	0
DB:	2	Gaps:	0
US-10-008-355-25 (1-10) x AC121125 (1-66762)			
OY 1 ThrClyclyAsnSerGlySerProValphe 10			
Db 57225 ACAGGAATAACAGCGGCTCTCCCTATTT 57254			
RESULT 16			
AC011338			
LOCUS	AC011338	120355 bp	DNA linear PRI 27-NOV-2001
DEFINITION	Homo sapiens chromosome 5 clone CTC-255N20, complete sequence.		
ACCESSION	AC011338		
VERSION	AC011338.8	GI:17105284	
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 120355)		
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 120355)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
REFERENCE	3 (bases 1 to 120355)		
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On Nov 27, 2001 this sequence version replaced gi:13752650. Draft Sequence Produced by DOE Joint Genome Institute www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu Quality: Phrap Quality >=40 99.9% of Sequence. Estimated Total Number of Errors is 0.1.		
FEATURES	Location/Qualifiers		
source	1. 120355		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="5"		
	/clone="CTC-255N20"		
BASE COUNT	37629 a 23902 c 23470 g 35354 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.13e+04	Length:	120355
Score:	45.00	Matches:	8
Percent Similarity:	88.89%	Conservative:	0
Best Local Similarity:	88.89%	Mismatches:	1
Query Match:	83.33%	Indels:	0
DB:	9	Gaps:	0
US-10-008-355-25 (1-10) x AC011338 (1-120355)			
OY 2 GlycylAsnSerGlySerProValphe 10			
Db 33504 GGAGGGAACACAGGGAGTCTGTGTTT 33530			
RESULT 17			
AC010251/c			
LOCUS	AC010251	175191 bp	DNA linear 16-APR-2000
DEFINITION	Homo sapiens chromosome 5 clone CTC-426D19, complete sequence.		
ACCESSION	AC010251		
VERSION	AC010251.3	GI:7109389	
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		

Not 100%

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 175191)  
 DOF Joint Genome Institute and Stanford Human Genome Center.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 175191)  
 DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 3 (bases 1 to 175191)  
 DOE Joint Genome Institute and Stanford Human Genome Center.  
 Direct Submission  
 Submitted (29-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 4 (bases 1 to 175191)  
 DOE Joint Genome Institute and Stanford Human Genome Center.  
 Direct Submission  
 Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 On Feb 29, 2000 this sequence version replaced gi:6600863.  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.5% of Sequence;  
 Estimated Total Number of Errors is 0.7.  
 Location/Qualifiers  
 1. 175191  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTC-426D19"  
 BASE COUNT 55384 a 33702 c 33905 g 52200 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.49e+04 Length: 175191  
 Score: 45.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 83.33% Indels: 0  
 DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AC010251 (1-175191)  
 Qy 2 GlyGlyAsnSerGlySerProvalPhe 10  
 Db 40572 GGAGGGAAACAGGGGAGTCTGTGTTT 40546

RESULT 18  
 AC116511  
 LOCUS  
 DEFINITION AC116511 194355 bp DNA linear HTG 08-JUL-2002  
 pieces.  
 AC116511  
 VERSION AC116511.3 GI:21703619  
 KEYWORDS HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 194355)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP24-314H15  
 Unpublished  
 2 (bases 1 to 194355)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhaltier, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Risse, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 194355)  
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhaltier, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Risse, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (08-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 8, 2002 this sequence version replaced gi:21700661.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence.submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L25486  
 Center clone name: 314\_H15  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 188139 bases at least Q40  
 Consensus quality: 190921 bases at least Q30  
 Consensus quality: 191755 bases at least Q20  
 Insert size: 196000; agarose-fp  
 Insert size: 192355; sum-of-contigs  
 Quality coverage: 7.0 in Q20 bases; agarose-fp  
 Quality coverage: 7.1 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
 \* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

```

1 1226: contig of 1226 bp in length
1227 1326: gap of 100 bp
1327 2632: contig of 1306 bp in length
2633 2732: gap of 100 bp
2733 4461: contig of 1729 bp in length
4462 4561: gap of 100 bp
4562 5781: contig of 1220 bp in length
5782 5881: gap of 100 bp
5882 7226: contig of 1345 bp in length
7227 7326: gap of 100 bp
7327 9815: contig of 2489 bp in length
9816 9915: gap of 100 bp
9916 13156: contig of 3241 bp in length
13157 13256: gap of 100 bp
13257 17362: contig of 4106 bp in length
17363 17462: gap of 100 bp
17463 21161: contig of 3699 bp in length
21162 21261: gap of 100 bp
21262 26310: contig of 5049 bp in length
26311 26410: gap of 100 bp
26411 28439: contig of 2029 bp in length
28440 28539: gap of 100 bp
28540 32012: contig of 3473 bp in length
32013 32112: gap of 100 bp
32113 38301: contig of 6189 bp in length
38302 38401: gap of 100 bp
38402 44818: contig of 6417 bp in length
44819 44918: gap of 100 bp
44919 51174: contig of 6256 bp in length
51175 51274: gap of 100 bp
51275 63416: contig of 12142 bp in length
63417 63516: gap of 100 bp
63517 76864: contig of 13348 bp in length
76865 76964: gap of 100 bp
76965 97902: contig of 20938 bp in length
97903 98002: gap of 100 bp
98003 117470: contig of 19468 bp in length
117471 117570: gap of 100 bp
117571 145371: contig of 27801 bp in length
145372 145471: gap of 100 bp
145472 194355: contig of 48884 bp in length.
  Location/Qualifiers
    1. .194355
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="RP24-314H15"
      /clone_lib="RPCI-24 Male Mouse BAC"
    1. .1326
      /note="assembly_fragment"
    1327. .2632
      /note="assembly_fragment"
    2733. .4461
      /note="assembly_fragment"
    4562. .5781
      /note="assembly_fragment"
    5882. .7226
      /note="assembly_fragment"
    7327. .9815
      /note="assembly_fragment"
    9916. .13156
      /note="assembly_fragment"
    13257. .17362
      /note="assembly_fragment"
    17463. .21161
      /note="assembly_fragment"
    21262. .26310
      /note="assembly_fragment"
    26411. .28439
      /note="assembly_fragment"
    28540. .32012
      /note="assembly_fragment"
  
```

# FEATURES

```

misc_feature      32113..38301
                  /note="assembly_fragment"
misc_feature      38402..44818
                  /note="assembly_fragment"
misc_feature      44919..51174
                  /note="assembly_fragment"
misc_feature      51275..63416
                  /note="assembly_fragment"
misc_feature      63517..76864
                  /note="assembly_fragment"
misc_feature      76965..97902
                  /note="assembly_fragment"
misc_feature      98003..117470
                  /note="assembly_fragment"
misc_feature      117571..145371
                  /note="assembly_fragment"
misc_feature      145472..194355
                  /note="assembly_fragment"
BASE COUNT      54684 a 42098 c 41214 g 54339 t 2020 others
ORIGIN
Alignment Scores:
  Pred. No.:      1.6e+04*
  Score:          45.00
  Percent Similarity: 90.00%
  Best Local Similarity: 80.00%
  Query Match:    83.33%
  DB:              2
US-10-008-355-25 (1-10) x AC116511 (1-194355)
QY      1 ThrGlycylAsnSerGlySerProValpHe 10
        |||||:|||||:|||||:|||||:|||||
Db 115302 ACAGGTGATCTTCAGGAAGTCCTGCCTTC 115331
RESULT 19
AC068065/c      195335 bp      DNA      linear      HTG 28-APR-2000
LOCUS          Mus musculus clone RP23-309J5 strain C57BL6/J, WORKING DRAFT
DEFINITION     SEQUENCE, 11 unordered pieces.
AC068065
AC068065.1 GI:76583325
VERSION         HTG: HTGS, PHASE1; HTGS_DRAFT.
KEYWORDS       Mus musculus.
SOURCE          Mus musculus
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS        Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
                Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
                Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,O.L., Maduro,V.B.,
                Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
                Stantripoop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
                Tionson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
                Wetherby,K.D. and Green,E.D.
TITLE          NISC Mouse Sequencing Initiative
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 195335)
AUTHORS        Green,E.D.
TITLE          Direct Submission
JOURNAL        Submitted (28-APR-2000) NIH Intramural Sequencing Center, 8717
                Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT       ----- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nihri.nih.gov
                ----- Project Information
                Center project name: yv
                Center clone name: 309J05
                ----- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
  
```







AUTHORS Glaser,P. and Kunst,F.  
 TITLE Listeria innocua, genome and applications  
 JOURNAL Patent: WO 0228891-A 4030 11-APR-2002;  
 Pasteur Institut (FR)

## FEATURES

source  
 1..349980  
 /organism="Listeria innocua"  
 /db\_xref="taxon:1642"  
 /note="seq 10 splitted into-seq 4029: from 300001 to  
 649980-seq 4030: from 600001 to 949980-seq 4031: from  
 900000 to 1163020"

BASE COUNT 116873 a 59062 c 69987 g 104055 t 3 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 2.47e+04 Length: 349980  
 Score: 45.00 Matches: 8  
 Percent Similarity: 90.00% Conservative: 1  
 Best Local Similarity: 80.00% Mismatches: 1  
 Query Match: 83.33% Indels: 0  
 DB: 6 Gaps: 0

US-10-008-355-25 (1-10) x AX417039 (1-349980)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10

||||| ||||||||| ||||||||| |||

Db 105352 ACCGGATCTAATTCGGTTCTCCGCTTTT 105381

## RESULT 22

AX417042 AX417042 349980 bp DNA linear PAT 14-JUN-2002  
 LOCUS Sequence 4033 from Patent WO0228891.  
 ACCESSION AX417042  
 VERSION AX417042.1 GI:21449652

## KEYWORDS

Listeria innocua.

## ORGANISM

Listeria innocua  
 Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

## REFERENCE

1  
 Glaser,P. and Kunst,F.  
 Listeria innocua, genome and applications  
 TITLE Patent: WO 0228891-A 4033 11-APR-2002;  
 JOURNAL Pasteur Institut (FR)

## FEATURES

Location/Qualifiers

1..349980  
 /organism="Listeria innocua"  
 /db\_xref="taxon:1642"

/note="seq 2058, original length: 3.011.208 replaced  
 by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to  
 0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:  
 0.900.001 to 1.249.980-seq 4035: 1.200.001 to  
 1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:  
 1.800.001 to 2.149.980-seq 4038: 2.100.001 to  
 2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:  
 2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"  
 BASE COUNT 117060 a 59094 c 70468 g 103358 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 2.47e+04 Length: 349980  
 Score: 45.00 Matches: 8  
 Percent Similarity: 90.00% Conservative: 1  
 Best Local Similarity: 80.00% Mismatches: 1  
 Query Match: 83.33% Indels: 0  
 DB: 6 Gaps: 0

US-10-008-355-25 (1-10) x AX417042 (1-349980)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10

||||| ||||||||| ||||||||| |||

Db 195617 ACCGGATCTAATTCGGTTCTCCGCTTTT 195646

## RESULT 23

## EFSFPREG

LOCUS E.faecalis sprE gene for serine proteinase homologue. linear BCT 22-DEC-1993  
 DEFINITION  
 ACCESSION \*Z12296  
 VERSION Z12296.1 GI:43337

## KEYWORDS

serine proteinase homologue; sprE gene.  
 ORGANISM Enterococcus faecalis.

## SOURCE

Bacteria; Firmicutes; Lactobacillales; Enterococcaceae;  
 Enterococcus.

## REFERENCE

1 (bases 1 to 1085)

## AUTHORS

Su,f.A. and Clewell,D.B.

## TITLE

A gene (sprE) downstream of gele of Enterococcus faecalis OGI-10  
 resembles serine proteinase determinant of Staphylococcus aureus  
 strain V8

## JOURNAL

Unpublished

## AUTHORS

Clewell,D.B.

## TITLE

Direct Submission

## JOURNAL

Submitted (07-JUN-1992) Don B Clewell, Biologic and Materials  
 Sciences, The University of, Michigan, School of Dentistry, 300 N.  
 Ingalls Bldg. Room 1198SE, Ann Arbor, Michigan, 48109-0402, USA

## FEATURES

Location/Qualifiers

1..1085  
 /organism="Enterococcus faecalis"

/strain="OGI-10"

/db\_xref="taxon:1351"

75..78

/note="Consensus"

91..945

/gene="sprE"

91..945

/gene="sprE"

/codon\_start=1

/transl\_table=11

/product="Staphylococcal serine proteinase homologue"

/protein\_id="CAA78168.1"

/db\_xref="GI:43338"

/db\_xref="SPTREMBL:Q47809"

/translation="MKFSIRKISAGFLFLVLTLAGFSLSANAEFIVPAESHRSRQ  
 KRSLDPEDRRQEVADTEAFPSAGISRIISPASKPGYISLCTGFVVGTVNTVNNHVA  
 ESFNKAVLNENAKDDAFYPRDGSATPFGRKVIDVAFSPNADIAVTVTKQNDRP  
 DGPELGELTFVLKFKESDTHVTISGPGKKNHTOWSHENDLFTSNFTDLENPLLF  
 YDIDTGGQSGSPIYNAQFEVVGVHSGGIRKQTGNHGQRLNEVNYNFIVNRVNEENK  
 RLSAVPA"

954..1008

/note="factor-independent"

terminator

BASE COUNT 364 a 189 c 223 g 309 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 513 Length: 1085  
 Score: 44.00 Matches: 7  
 Percent Similarity: 90.00% Conservative: 2  
 Best Local Similarity: 70.00% Mismatches: 1  
 Query Match: 81.48% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x EFSFPREG (1-1085)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10

||||| ||||||||| ||||||||| |||

Db 760 ACAGGTGGTCAATCTGGTTCCTCCCAATCTAT 789

## RESULT 24

AF008576/c

## LOCUS

AF008576

DEFINITION Homo sapiens galanin receptor gene, 5'UTR.

ACCESSION AF008576

VERSION AF008576.1 GI:2352810

KEYWORDS Homo sapiens.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

linear PRI 05-FEB-1998

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1912)

AUTHORS Lorimer,D.D., Matkowskyj,K. and Benya,R.V.

TITLE Cloning, chromosomal location, and transcriptional regulation of

JOURNAL the human galanin-1 receptor gene (GALN1R)

MEDLINE 98086390

PUBMED 9425310

REFERENCE 2 (bases 1 to 1912)

AUTHORS Lorimer,D.D., Lui,W., Matkowskyj,K. and Benya,R.V.

TITLE Direct Submission

JOURNAL Submitted (16-JUN-1997) Medicine, University of Illinois at

FEATURES Chicago, 840 S. Wood St. (M/C 787), Chicago, IL 60612, USA

Location/Qualifiers

1. .1912

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="15"

/map="15q24"

1. .1912

/note="galanin receptor gene"

BASE COUNT 460 a 498 c 573 g 381 t

ORIGIN

Alignment Scores:

Pred. No.: 777 Length: 1912

Score: 44.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 81.48% Indels: 0

DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AF008576 (1-1912)

Oy 1 ThrGlyGlyAsnSerGlySerPro 8

|||||

Db 1210 ACTGGTGGAAACTCGGGCTCCCC 1187

|||||

RESULT 25

BC019638

LOCUS 2385 bp mRNA linear ROD 07-AUG-2002

DEFINITION Mus musculus, Similar to RIKEN cDNA 4632417K18 gene, clone

ACCESSION BC019638

VERSION BC019638

KEYWORDS BC019638.1 GI:18044022

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE Strausberg,R.

JOURNAL Direct Submission

Submitted (19-DEC-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowls, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 40 Row: p Column: 1.

Location/Qualifiers

1. .2385

/organism="Mus musculus"

/db\_xref="taxon:10090"

/map="CZECH II"

/clone="IMAGE:4010556"

/tissue\_type="Mammary tumor metastasized to lung.

MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR

enhancer."

/clone\_lib="NCI\_CGAP\_Lu30"

/lab\_host="DH10B"

/note="Vector: pCMV-SPORT6"

<1. .1197

/codon\_start=1

/product="Similar to RIKEN cDNA 4632417K18 gene"

/protein\_id="AAH19638.1"

/db\_xref="GI:18044023"

/translation="IRDTLRKGRFCTFTESDDWKLIGNLSIIETQPVDELEGKLF

QVAAELPKNPVAVSTVONGSENENFHKLEEVIMNEYTLKKEGKLRAIKKSEKR

KKKASLFKVKHKEFGKTRNSTPVQVVKHLSRVSDSVGLMNNNGNAGCATCFVKE

LYTLTQHVIVASIVGEGIDSSSEWAIISQCVKTFDYELLPTGDKFFMKVPWFESD

KHLDYAVLELKENGQVPAGLYHRIKVPVPSGLIYIIGHPEGKKSIDCCCTVVPQSSR

RKKQENFQAREAGCFSTSFIMVYTORSEFQMLHNSDVVYDTYDFFGSGSPVFD

SNGLVAMHAAGITCTYQAGVSNIIIEFGSINESIDDDHMKQDKKKEYNTISGNVNE

MLSIDF"

BASE COUNT 764 a 413 c 522 g 686 t

ORIGIN

Alignment Scores:

Pred. No.: 914 Length: 2385

Score: 44.00 Matches: 8

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 88.89% Mismatches: 0

Query Match: 81.48% Indels: 0

DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x BC019638 (1-2385)

Oy 2 GlyGlyAsnSerGlySerProValpHe 10

|||||

Db 973 GGTGGTCTCTGGATCCCCAGTATTT 999

|||||

RESULT 26

AC017307

LOCUS 36547 bp DNA linear HTG 09-DEC-1999

DEFINITION AC017307 Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered

pieces.

ACCESSION AC017307

VERSION AC017307.1 GI:6553679

KEYWORDS HTG: HTGS\_PHASE2.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 36547)

Adams,M. and Venter,J.C.

Direct Submission

Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

This sequence was identified as CDM:10210224 by the submitter.

For more information on this record e-mail to [fly@celera.com](mailto:fly@celera.com).

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

Location/Qualifiers

1. .36547

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

FEATURES

source

```

BASE COUNT 10301 a 8119 c 7889 g 10238 t
ORIGIN
Alignment Scores:
Pred. No.: 6.75e+03 36547
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AC017307 (1-36547)
Qy 2 GlyGlyAsnSerGlySerProValphe 10
|||||:|||||:|||||:|||||
Db 32524 GGGGGGAGTTCGGGGAGCCAGTATT 32550
AD000002.1 GI:1702966

RESULT 27
MSGY154/c 40221 bp DNA linear BCT 03-DEC-1996
LOCUS Mycobacterium tuberculosis sequence from clone y154.
DEFINITION Mycobacterium tuberculosis (clone: y154) ds-DNA.
ACCESSION AB000002
VERSION AD000002.1
KEYWORDS Mycobacterium tuberculosis
SOURCE Mycobacterium tuberculosis (clone: y154) ds-DNA.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 40221)
REFERENCE
AUTHORS Du, L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT GSDB:S:1004707
FEATURES
source
1..40221
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
/clone="y154"

BASE COUNT 6760 a 12728 c 13608 g 7094 t 31 others
ORIGIN
Alignment Scores:
Pred. No.: 7.24e+03 40221
Score: 44.00 Matches: 8
Percent Similarity: 80.00% Conservative: 0
Best Local Similarity: 80.00% Mismatches: 2
Query Match: 81.48% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x MSGY154 (1-40221)
Qy 1 ThrGlyGlyAsnSerGlySerProValphe 10
|||||:|||||:|||||:|||||
Db 10876 ACCGGTGGCAACGGCGGTTCACCACTTTC 10847
AC100192

RESULT 28
AC100192 58722 bp DNA linear HTG 22-NOV-2001
LOCUS Mus musculus clone RP23-56N11, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
ACCESSION AC100192
VERSION AC100192.1 GI:17047558
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 58722)
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-56N11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 58722)

```

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Melidrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## COMMENT

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Project Information

Center project name: L14286

Center clone name: 56\_N11

\* NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
\* 695 794: contig of 694 bp in length  
\* 795 1428: contig of 634 bp in length  
\* 1429 1528: gap of 100 bp  
\* 1529 2207: contig of 679 bp in length  
\* 2208 2307: gap of 100 bp  
\* 2308 3002: contig of 695 bp in length  
\* 3003 3102: gap of 100 bp  
\* 3103 3808: contig of 706 bp in length  
\* 3809 3908: gap of 100 bp  
\* 3909 4595: contig of 687 bp in length  
\* 4596 4695: gap of 100 bp  
\* 4696 5398: contig of 703 bp in length  
\* 5399 5498: gap of 100 bp  
\* 5499 6189: contig of 691 bp in length  
\* 6190 6289: gap of 100 bp  
\* 6290 6987: contig of 698 bp in length  
\* 6988 7087: gap of 100 bp  
\* 7088 7784: contig of 697 bp in length  
\* 7785 7884: gap of 100 bp  
\* 7885 8560: contig of 676 bp in length  
\* 8561 8660: gap of 100 bp  
\* 8661 9365: contig of 705 bp in length  
\* 9366 9465: gap of 100 bp  
\* 9466 10156: contig of 691 bp in length

\* 10157 10256: gap of 100 bp in length  
\* 10257 10939: contig of 683 bp in length  
\* 10940 11039: gap of 100 bp in length  
\* 11040 11671: contig of 632 bp in length  
\* 11672 11771: gap of 100 bp in length  
\* 11772 12491: contig of 720 bp in length  
\* 12492 12591: gap of 100 bp in length  
\* 12592 13313: contig of 722 bp in length  
\* 13314 13413: gap of 100 bp in length  
\* 13414 14119: contig of 706 bp in length  
\* 14120 14219: gap of 100 bp in length  
\* 14220 14917: contig of 698 bp in length  
\* 14918 15017: gap of 100 bp in length  
\* 15018 15711: contig of 694 bp in length  
\* 15712 15811: gap of 100 bp in length  
\* 15812 16491: contig of 680 bp in length  
\* 16492 16591: gap of 100 bp in length  
\* 16592 17283: contig of 692 bp in length  
\* 17284 17383: gap of 100 bp in length  
\* 17384 18078: contig of 695 bp in length  
\* 18079 18178: gap of 100 bp in length  
\* 18179 18875: contig of 697 bp in length  
\* 18876 18975: gap of 100 bp in length  
\* 18976 19670: contig of 695 bp in length  
\* 19671 19770: gap of 100 bp in length  
\* 19771 20487: contig of 717 bp in length  
\* 20488 20587: gap of 100 bp in length  
\* 20588 21288: contig of 701 bp in length  
\* 21289 21388: gap of 100 bp in length  
\* 21389 22100: contig of 712 bp in length  
\* 22101 22200: gap of 100 bp in length  
\* 22201 22896: contig of 696 bp in length  
\* 22897 22996: gap of 100 bp in length  
\* 22997 23680: contig of 684 bp in length  
\* 23681 23780: gap of 100 bp in length  
\* 23781 24474: contig of 694 bp in length  
\* 24475 24574: gap of 100 bp in length  
\* 24575 25265: contig of 691 bp in length  
\* 25266 25365: gap of 100 bp in length  
\* 25366 26075: contig of 710 bp in length  
\* 26076 26175: gap of 100 bp in length  
\* 26176 26880: contig of 705 bp in length  
\* 26881 26980: gap of 100 bp in length  
\* 26981 27696: contig of 716 bp in length  
\* 27697 27796: gap of 100 bp in length  
\* 27797 28485: contig of 689 bp in length  
\* 28486 28585: gap of 100 bp in length  
\* 28586 29276: contig of 691 bp in length  
\* 29277 29376: gap of 100 bp in length  
\* 29377 30066: contig of 690 bp in length  
\* 30067 30166: gap of 100 bp in length  
\* 30167 30867: contig of 701 bp in length  
\* 30868 30967: gap of 100 bp in length  
\* 30968 31658: contig of 691 bp in length  
\* 31659 31758: gap of 100 bp in length  
\* 31759 32459: contig of 701 bp in length  
\* 32460 32559: gap of 100 bp in length  
\* 32560 33260: contig of 701 bp in length  
\* 33261 33360: gap of 100 bp in length  
\* 33361 34063: contig of 703 bp in length  
\* 34064 34163: gap of 100 bp in length  
\* 34164 34851: contig of 688 bp in length  
\* 34852 34951: gap of 100 bp in length  
\* 34952 35643: contig of 692 bp in length  
\* 35644 35743: gap of 100 bp in length  
\* 35744 36443: contig of 699 bp in length  
\* 36443 36542: gap of 100 bp in length  
\* 36543 37236: contig of 694 bp in length  
\* 37237 37336: gap of 100 bp in length  
\* 37337 38033: contig of 697 bp in length  
\* 38034 38133: gap of 100 bp in length  
\* 38134 38822: contig of 689 bp in length  
\* 38823 38922: gap of 100 bp in length

\* 38923 39601: contig of 679 bp in length  
\* 39602 39701: gap of 100 bp in length  
\* 39702 40385: contig of 684 bp in length  
\* 40386 40485: gap of 100 bp in length  
\* 40486 41183: contig of 698 bp in length  
\* 41184 41283: gap of 100 bp in length  
\* 41284 41983: contig of 700 bp in length  
\* 41984 42083: gap of 100 bp in length  
\* 42084 42767: contig of 684 bp in length  
\* 42768 42867: gap of 100 bp in length  
\* 42868 43561: contig of 694 bp in length  
\* 43562 43661: gap of 100 bp in length  
\* 43662 44360: contig of 699 bp in length  
\* 44361 44460: gap of 100 bp in length  
\* 44461 45156: contig of 696 bp in length  
\* 45157 45256: gap of 100 bp in length  
\* 45257 45958: contig of 702 bp in length  
\* 45959 46058: gap of 100 bp in length  
\* 46059 46768: contig of 710 bp in length  
\* 46769 46868: gap of 100 bp in length  
\* 46869 47562: contig of 694 bp in length  
\* 47563 47662: gap of 100 bp in length  
\* 47663 48365: contig of 703 bp in length  
\* 48366 48465: gap of 100 bp in length  
\* 48466 49164: contig of 699 bp in length  
\* 49165 49264: gap of 100 bp in length  
\* 49265 49960: contig of 696 bp in length  
\* 49961 50060: gap of 100 bp in length  
\* 50061 50751: contig of 691 bp in length  
\* 50752 50851: gap of 100 bp in length  
\* 50852 51565: contig of 714 bp in length  
\* 51566 51665: gap of 100 bp in length  
\* 51666 52359: contig of 694 bp in length  
\* 52360 52459: gap of 100 bp in length  
\* 52460 53151: contig of 692 bp in length  
\* 53152 53251: gap of 100 bp in length  
\* 53252 53957: contig of 706 bp in length  
\* 53958 54057: gap of 100 bp in length  
\* 54058 54762: contig of 705 bp in length  
\* 54763 54862: gap of 100 bp in length

Alignment Scores:

Pred. No.: 9.55e+03 Length: 58722  
Score: 44.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 81.48% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AC100192 (1-58722)

Oy 2 GlyGlyAsnSerGlySerProValphe 10

Db 31251 GGGGAGAGATTCTGGAGTCCTGTATT 31277  
|||||:|||||:|||||:|||||

RESULT 29

AC119569\_2/c

WPCOMMENT

Sequence split into 4 fragments LOCUS AC119569 Accession AC119569

Fragment Name	Begin	End
AC119569_0	1	110000
AC119569_1	100001	210000
AC119569_2	200001	310000
AC119569_3	300001	397831

Continuation (3 of 4) of AC119569 from base 200001 (AC119569 Homo sapiens clone RP13-

Alignment Scores:

Pred. No.: 1.51e+04 Length: 110000  
Score: 44.00 Matches: 7  
Percent Similarity: 90.00% Conservatives: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 81.48% Indels: 0  
DB: 2 Gaps: 0

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

```

/notes="match to EST BE172154 (NID:g8634880)"
6310..6342
/rpt_family="L1"
6343..6534
/rpt_family="MER2_type"
6535..6830
/rpt_family="Alu"
6831..6948
/rpt_family="MER2_type"
6949..7453
/rpt_family="L1"
7522..7815
/rpt_family="Alu"
8292..8347
/rpt_family="MIR"
8381..8688
/rpt_family="Alu"
8827..9065
/rpt_family="MIR"
10185..10227
/rpt_family="ERV1"
10229..10336
/rpt_family="MIR"
10337..10648
/rpt_family="Alu"
10649..10661
/rpt_family="MIR"
10662..10966
/rpt_family="Alu"
10967..11016
/rpt_family="MIR"
11533..11838
/rpt_family="Alu"
12642..13853
/notes="CpG island (GC=73.3, o/e=0.90, #CpGs=137)"
13062..13216
/notes="match to EST BE792154 (NID:g10213352)"
13268..13339
/rpt_family="(CCCG)n"
14379..14373
/rpt_family="L2"
14552..15119
/rpt_family="Achobo"
15257..15567
/rpt_family="Alu"
15645..15745
/rpt_family="MIR"
15849..15930
/rpt_family="GA-rich"
16260..16429
/rpt_family="MIR"
17155..17455
/rpt_family="Alu"
17462..17607
/rpt_family="MIR"
17656..17864
/rpt_family="MER1_type"
17867..17961
/rpt_family="L2"
18268..18395
/rpt_family="Alu"
19184..19456
/rpt_family="Alu"

```

## Alignment Scores:

```

Pred. No.: 2.01e+04 Length: 162609
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 9 Gaps: 0

```

US-10-008-355-25 (1-10) x AC108022 (1-162609)

```

Qy 1 ThrGlyGlyAsnSerGlySerPro 8
AC009741/c
Db 42377 ACTGCTGGAATCTCGCATCCCCA 42400
|||||
|||||
|||||
|||||
|||||

```

## RESULT 31

AC009741/c

## LOCUS

## DEFINITION

Drosophila melanogaster, chromosome 3R, region 87B-87B, BAC clone

BACR44K17, complete sequence.

## ACCESSION

AC009741

## VERSION

AC009741.5 GI:13096041

## KEYWORDS

HTG.

## SOURCE

Drosophila melanogaster.

## ORGANISM

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 166863)

## AUTHORS

Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,

Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,

Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,

Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,

Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,

Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,

Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,

McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,

Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,

Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Schaefer,F.,

Staverton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,

Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Sequencing of Drosophila chromosome 3R, region 87B-87B

Unpublished

## TITLE

2 (bases 1 to 166863)

## JOURNAL

Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

Rubin,G.M.

## TITLE

Submitted (30-AUG-1999) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

## JOURNAL

On Feb 22, 2001 this sequence version replaced gi:5912616.

## COMMENT

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome

shotgun and from subclones of this BAC and its neighboring clones.

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send emailto [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

## FEATURES

Location/Qualifiers

1..166863

/Organism="Drosophila melanogaster"

/Strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="3R"

/map="87B-87B"

/Clone="BACR44K17 (D976)"

/Clone\_lib="RPC1-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

pBACE3.6)"

BASE COUNT 47964 a 36383 c 35720 g 46796 t

ORIGIN

Alignment Scores:



Pred. No.: 2.05e+04 Length: 166863  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 3 Gaps: 0

US-10-008-355-25 (1-10) x AC009741 (1-166863)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10

|||||:|||||:|||||:|||||

Db 9520 GGGGGCAGTTCGGGAGCCAGCAGTATT 9494

RESULT 32

AC108085/c

LOCUS

DEFINITION Homo sapiens chromosome 5 clone CTD-2210F7, WORKING DRAFT SEQUENCE,  
 6 unordered pieces.

ACCESSION AC108085

VERSION AC108085.1 GI:18369931

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 166978)

DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 166978)

DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

-----

Project Information

Center Project Name: 708316

Center clone name: CITB-H1\_2210F7

-----

Summary Statistics

Consensus quality: 163223 bases at least Q40

Consensus quality: 164704 bases at least Q30

Consensus quality: 165029 bases at least Q20

Estimated insert size: 179000; agarose-fp estimation

Estimated insert size: 166478; sum-of-contigs estimation

Quality coverage: 8.17 in Q20 bases; agarose-fp estimation

Quality coverage: 8.78 in Q20 bases; sum-of-contigs estimation

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 6 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1111: contig of 1111 bp in length

\* 1211: gap of unknown length

\* 1212: contig of 1320 bp in length

\* 2531: gap of unknown length

\* 2532: contig of 1453 bp in length

\* 2632: gap of unknown length

\* 4085: contig of 46472 bp in length

\* 4185: gap of unknown length

\* 50657: contig of 50460 bp in length

\* 50757: contig of 50460 bp in length

\* 101217: contig of 50460 bp in length

\* 101317: contig of 65662 bp in length.

\* Location/Qualifiers

1. 166978

/organism="Homo sapiens"

FEATURES

source

/db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2210F7"  
 /clone\_lib="CalTech human BAC library D"  
 BASE COUNT 53321 a 32702 c 32339 g 48053 t 563 others  
 ORIGIN

Alignment Scores:

Pred. No.: 2.05e+04 Length: 166978  
 Score: 44.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 81.48% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AC108085 (1-166978)

Qy 1 ThrGlyGlyAsnSerGlySerProVal 9

|||||:|||||:|||||:|||||

Db 2529 ACAGGGGGCAATTCGGGGGCCCCGTT 2503

RESULT 33

AC036185

LOCUS

DEFINITION Homo sapiens chromosome 4 clone RP11-67986 map 4, WORKING DRAFT

SEQUENCE, 15 unordered pieces.

ACCESSION AC036185

VERSION AC036185.2 GI:8576273

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 171279)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 4, clone RP11-67986

Unpublished

2 (bases 1 to 171279)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,K., Dearellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 21, 2000 this sequence version replaced gi:7523854.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: U9247

Center clone name: 679\_B\_6  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 163254 bases at least Q40  
 Consensus quality: 166954 bases at least Q30  
 Consensus quality: 168680 bases at least Q20  
 Insert size: 170000; agarose-fp  
 Quality coverage: 7.1 in Q20 bases; agarose-fp  
 Quality coverage: 7.1 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 15 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1385: contig of 1385 bp in length  
 \* 1386 1485: gap of 100 bp  
 \* 1486 2835: contig of 1350 bp in length  
 \* 2836 2935: gap of 100 bp  
 \* 2936 4108: contig of 1173 bp in length  
 \* 4109 4208: gap of 100 bp  
 \* 4209 7507: contig of 3299 bp in length  
 \* 7508 7607: gap of 100 bp  
 \* 7608 13869: contig of 6262 bp in length  
 \* 13870 13969: gap of 100 bp  
 \* 13970 20810: contig of 6841 bp in length  
 \* 20811 20910: gap of 100 bp  
 \* 20911 28893: contig of 7983 bp in length  
 \* 28894 28993: gap of 100 bp  
 \* 28994 33086: contig of 10093 bp in length  
 \* 33087 39186: gap of 100 bp  
 \* 39187 50077: contig of 10891 bp in length  
 \* 50078 50177: gap of 100 bp  
 \* 50178 61526: contig of 11349 bp in length  
 \* 61527 61626: gap of 100 bp  
 \* 61627 75529: contig of 13903 bp in length  
 \* 75530 75623: gap of 100 bp  
 \* 75630 91544: contig of 15915 bp in length  
 \* 91545 91644: gap of 100 bp  
 \* 91645 108750: contig of 17106 bp in length  
 \* 108751 108850: gap of 100 bp  
 \* 108851 136877: contig of 28027 bp in length  
 \* 136878 136977: gap of 100 bp  
 \* 136978 171279: contig of 34302 bp in length.  
 Location/Qualifiers  
 1. 171279

FEATURES  
source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-679B6"  
 /clone.lib="RPC1-11 Human Male BAC"  
 1. 1385  
 /note="assembly\_fragment"  
 1486..2835  
 /note="assembly\_fragment"  
 2936..4108  
 /note="assembly\_fragment"  
 4209..7507  
 /note="assembly\_fragment"  
 7608..13869  
 /note="assembly\_fragment"  
 13970..20810  
 /note="assembly\_fragment"  
 20911..28893  
 /note="assembly\_fragment"  
 28994..33086

/note="assembly\_fragment"  
 39187..50077  
 /note="assembly\_fragment"  
 50178..61526  
 /note="assembly\_fragment"  
 61627..75529  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:right  
 75630..91544  
 /note="assembly\_fragment"  
 91645..108750  
 /note="assembly\_fragment"  
 108851..136877  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:right  
 136978..171279  
 /note="assembly\_fragment"  
 BASE COUNT 48027 a 36049 c 36303 g 49499 t 1401 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.09e+04 Length: 171279  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AC036185 (1-171279)

Qy 1 ThrGlyGlyAsnSerGlySerPro 8

Db 126420 ACTGTGGAAATTCGTGATCCCA 126443

RESULT 34

AC068590

LOCUS

DEFINITION

Homo sapiens chromosome 8 clone RP11-131K16 map 8, WORKING DRAFT

SEQUENCE, 30 unordered pieces.

AC068590

VERSION

AC068590.2 GI:8389591

KEYWORDS

HTG: HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 174253)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 8, clone RP11-131K16

Unpublished

2 (bases 1 to 174253)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGuire,A., McKernan,K., McPheters,R.,

Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,

linear HTG 24-AUG-2002



```

misc_feature 18436..19919
/note="assembly_fragment
clone_end=T7
vector_side:right"
20020..22351
/note="assembly_fragment"
22452..25473
/note="assembly_fragment"
25574..28192
/note="assembly_fragment"
28293..32574
/note="assembly_fragment"
32675..36983
/note="assembly_fragment"
37084..41557
/note="assembly_fragment"
41658..46775
/note="assembly_fragment"
46876..52390

Alignment Scores:
Pred. No.: 2.12e+04 Length: 174253
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AC068590 (1-174253)
Qy 1 ThrGlyGlyAsnSerglySerPro 8
Db 40541 ACAGGAGGACAGTGCAGCCCT 40564

RESULT 35
AC007724/c
LOCUS AC007724 178199 bp DNA linear INV 17-FEB-2001
DEFINITION Drosophila melanogaster, chromosome 3R, region 87B-87B, BAC clone
BACR30N15, complete sequence.
AC007724
AC007724.4 GI:12957626
HTG.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 178199)
Celinker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Ananidis,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garb,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Idegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 87B-87B
Unpublished
2 (bases 1 to 178199)
Celinker,S.E., Agbanyani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.

TITLE Direct Submission
JOURNAL Submitted (04-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Feb 17, 2001 this sequence version replaced gi:5670544.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
FEATURES
Location/Qualifiers
1..178199
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="87B-87B"
/clone="BACR30N15 (D693)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACe3.6)"
BASE COUNT 50162 a 39188 c 39108 g 49741 t
ORIGIN
Alignment Scores:
Pred. No.: 2.15e+04 Length: 178199
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 3 Gaps: 0

US-10-008-355-25 (1-10) x AC007724 (1-178199)
Qy 2 GlyGlyAsnSerglySerProValpHe 10
Db 101032 GCGGGAGTTCGGGGAGCCCATTTT 101006

RESULT 36
AC073533/c
LOCUS AC073533 182056 bp DNA linear PRI 07-MAR-2002
DEFINITION Homo sapiens Xp BAC RP11-589J20 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
AC073533
AC073533.19 GI:14670054
HTG.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 182056)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blimie,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burckett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotiz,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Franz,C., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havliak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsif., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

```

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Lewis, D.C., Lewis, L., Li, J., Lichte, O., Lieu, C., Liu, J., Liu, W., Loiseleg, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., and Gibbs, R.

# TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Direct Submission  
Unpublished  
2 (bases 1 to 182056)  
Worley, K.C.

Direct Submission  
Submitted (22-JUN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 182056)  
Worley, K.C.

Direct Submission  
Submitted (11-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 182056)  
Worley, K.C.

Direct Submission  
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2001 this sequence version replaced gi:13876434.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

## ANNOTATION OF FEATURES:

SPFS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality

standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: [http://gc.bcm.tmc.edu:8088/quality\\_info/genbank.annotation.html](http://gc.bcm.tmc.edu:8088/quality_info/genbank.annotation.html).

## QUALSTAT-REPORT-----

----- Summary Statistics -----  
Contig length: 106921  
Phrap values in estimate: 106537  
Average error rate (BCM-Phrap estimate): 3.16422e-05  
Fraction of Phrap values less than 40 : 0.0050405  
Number of consensus changing edits: 0  
Number of N's in consensus : 0

----- Consensus changing edits -----  
Position Original+Context Edited+Context

----- Distribution of Quality < 40 Bases -----

500	5	10	15	20	25	30	35	40
450	Phrap Value Range							
400								
350								
300								
250								
200								
150								
100								
50								
0								

Version: 1.01 qxfo.

FEATURES	Location/Qualifiers
source	1. 182056
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/clone="RP11-589J20"
repeat_region	3077..3115
repeat_region	/rpt_family="(A)n"
repeat_region	3329..4073
repeat_region	/rpt_family="L1PB3"
repeat_region	4074..4095
repeat_region	/rpt_family="AT-rich"
repeat_region	complement(4120..4491)
repeat_region	/rpt_family="L1MB7"
repeat_region	4594..4662
repeat_region	/rpt_family="L1PB3"
repeat_region	complement(5037..5323)
repeat_region	/rpt_family="AluY"
repeat_region	complement(5446..5779)
repeat_region	/rpt_family="L1MB7"
repeat_region	8364..8480
repeat_region	/rpt_family="AluJb"
repeat_region	8481..8504
repeat_region	/rpt_family="(TAAA)n"
repeat_region	8547..8834
repeat_region	/rpt_family="AluSx"
repeat_region	8835..8865
repeat_region	/rpt_family="AT-rich"
repeat_region	complement(9275..9321)
repeat_region	/rpt_family="L1MA7"
repeat_region	10355..10529

```

repeat_region      /rpt_family="MER3"
10679. .10745
/rpt_family="(CA)n"
complement(11417. .11637)
/rpt_family="MER58A"
repeat_region      complement(11762. .12128)
/rpt_family="THE1C"
12885. .13502
/rpt_family="LTR14B"
repeat_region      13592. .13895
/rpt_family="AluX"
repeat_region      14471. .15062
/rpt_family="LTR26"
repeat_region      complement(15204. .15372)
/rpt_family="MER20"
repeat_region      15462. .15486
/rpt_family="(TG)n"
repeat_region      15702. .15851
/rpt_family="FRAM"
repeat_region      15942. .16069
/rpt_family="Tigger3(Golem)"
repeat_region      16067. .16323
/rpt_family="Tigger3(Golem)"
repeat_region      complement(17669. .17731)
/rpt_family="L2"

```

## Alignment Scores:

```

Pred. No.:      2.19e+04      Length:      182056
Score:          44.00      Matches:      7
Percent Similarity: 90.00%      Conservative: 2
Best Local Similarity: 70.00%      Mismatches: 1
Query Match:      81.48%      Indels: 0
DB:              9      Gaps: 0

```

US-10-008-355-25 (1-10) x AC073533 (1-182056)

Qy 1 ThrGlyGlyAsnSerGlySerProValphe 10

```

|||||
Db 40810 ACTGGGGCAACTCAGCAAGTCCTATTAT 40781

```

## RESULT 37

```

AC124514      AC124514      187036 bp      DNA      linear      HTG 14-JUN-2002
LOCUS      Mus musculus chromosome UNK clone RP23-450H18, WORKING DRAFT
DEFINITION      SQUENCE, 14 unordered pieces.

```

ACCESSION AC124514

VERSION AC124514.1 GI:21426636

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 187036)

McPherson,J.D. and Waterston,R.H.

Unpublished

2 (bases 1 to 187036)

McPherson,J.D. and Waterston,R.H.

Direct Submission

Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

## COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0450H18
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%

```

Chemistry: Dye-primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 179940 bases at least Q40  
 Consensus quality: 181332 bases at least Q30  
 Consensus quality: 182301 bases at least Q20  
 Insert size: 184000; agarose-fp  
 Insert size: 188324; sum-of-contigs  
 Quality coverage: 6.94 in Q20 bases; agarose-fp  
 Quality coverage: 6.21 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 14 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 1135: contig of 1135 bp in length
* 1136 1235: gap of unknown length
* 1236 2437: contig of 1202 bp in length
* 2438 2537: gap of unknown length
* 2538 4634: contig of 2097 bp in length
* 4635 4734: gap of unknown length
* 4735 7577: contig of 2743 bp in length
* 7578 11704: gap of unknown length
* 11705 11804: contig of 4127 bp in length
* 11805 15430: gap of unknown length
* 15431 15530: contig of 3626 bp in length
* 15531 26140: gap of unknown length
* 26141 26240: contig of 10610 bp in length
* 26241 38458: gap of unknown length
* 38459 38558: contig of 12218 bp in length
* 38559 56497: gap of unknown length
* 56498 56597: contig of 17939 bp in length
* 56598 74135: gap of unknown length
* 74136 74235: contig of 17538 bp in length
* 74236 98938: gap of unknown length
* 98939 99039: contig of 24703 bp in length
* 99039 121242: gap of unknown length
* 121243 121343: contig of 22204 bp in length
* 121343 146821: gap of unknown length
* 146822 146921: contig of 25479 bp in length
* 146922 187036: gap of unknown length
* 187036 187036: contig of 40115 bp in length.

```

## FEATURES

## source

```

1..187036
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-450H18"
misc_feature      1..1135
/note="assembly_name:Contig8"
misc_feature      1236..2437
/note="assembly_name:Contig20"
misc_feature      2538..4634
/note="assembly_name:Contig21"
misc_feature      4735..7477
/note="assembly_name:Contig22"
misc_feature      7578..11704
/note="assembly_name:Contig23"
misc_feature      11805..15430
/note="assembly_name:Contig24"
vector_side:left"
15531..26140
/note="assembly_name:Contig25"
26241..38458
/note="assembly_name:Contig26"
38559..56497
/note="assembly_name:Contig27"
56598..74135

```

```

/note="assembly_name:Contig28
clone_end:77
vector_side:left"
misc_feature
74236..98938
/note="assembly_name:Contig29"
99039..121242
/note="assembly_name:Contig30"
121343..146821
/note="assembly_name:Contig31"
146922..187036
/note="assembly_name:Contig32"
55566 a 36844 c 36962 g 56345 t 1319 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.23e+04 Length: 187036
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0

```

US-10-008-355-25 (1-10) x AC124514 (1-187036)

```

QY 2 GlycylAsnSerGlySerProValpHe 10
Db 65766 GGTGGTGCTCTCGATCCAGTATT 65792

```

#### RESULT 38

```

AC051649
LOCUS AC051649 189014 bp DNA linear HTG 16-AUG-2002
DEFINITION Homo sapiens chromosome 11 clone RP11-534I22 map 11, *** SEQUENCING
IN PROGRESS ***, 6 unordered pieces.
ACCESSION AC051649
VERSION 9 GI:22267727
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEPIN.
SOURCE human.

```

```

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 189014)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
TITLE Homo sapiens chromosome 11, clone RP11-534I22

```

```

REFERENCE
2 (bases 1 to 189014)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castelle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,J.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

```

3 (bases 1 to 189014)
REFERENCE
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camara,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,

```

```

TITLE
JOURNAL
COMMENT
On Aug 16, 2002 this sequence version replaced gi:22165331.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9681
Center clone name: 534_1_22
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 55890: contig of 55890 bp in length
* 55991 55990: gap of 100 bp
* 55991 58431: contig of 2441 bp in length
* 58432 58531: gap of 100 bp
* 58532 75875: contig of 17344 bp in length
* 75876 75975: gap of 100 bp
* 75976 80404: contig of 4429 bp in length
* 80405 80504: gap of 100 bp
* 80505 164904: contig of 84400 bp in length
* 164905 165004: gap of 100 bp
* 165005 189014: contig of 24010 bp in length.
FEATURES
Location/Qualifiers
source
1. 189014
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="RP11-534I22"
/clone_lib="RP11-534I22"
/clone_lib="RP11-534I22"
Human Male BAC"
BASE COUNT 37986 a 57251 c 54334 g 38738 t 665 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.25e+04 Length: 189014
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0
US-10-008-355-25 (1-10) x AC051649 (1-189014)
QY 1 ThrGlyAsnSerGlySerPro 8
|||||

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 55890: contig of 55890 bp in length
* 55991 55990: gap of 100 bp
* 55991 58431: contig of 2441 bp in length
* 58432 58531: gap of 100 bp
* 58532 75875: contig of 17344 bp in length
* 75876 75975: gap of 100 bp
* 75976 80404: contig of 4429 bp in length
* 80405 80504: gap of 100 bp
* 80505 164904: contig of 84400 bp in length
* 164905 165004: gap of 100 bp
* 165005 189014: contig of 24010 bp in length.
FEATURES
Location/Qualifiers
source
1. 189014
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="RP11-534I22"
/clone_lib="RP11-534I22"
Human Male BAC"
BASE COUNT 37986 a 57251 c 54334 g 38738 t 665 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.25e+04 Length: 189014
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0
US-10-008-355-25 (1-10) x AC051649 (1-189014)
QY 1 ThrGlyAsnSerGlySerPro 8
|||||

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 55890: contig of 55890 bp in length
* 55991 55990: gap of 100 bp
* 55991 58431: contig of 2441 bp in length
* 58432 58531: gap of 100 bp
* 58532 75875: contig of 17344 bp in length
* 75876 75975: gap of 100 bp
* 75976 80404: contig of 4429 bp in length
* 80405 80504: gap of 100 bp
* 80505 164904: contig of 84400 bp in length
* 164905 165004: gap of 100 bp
* 165005 189014: contig of 24010 bp in length.
FEATURES
Location/Qualifiers
source
1. 189014
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="RP11-534I22"
/clone_lib="RP11-534I22"
Human Male BAC"
BASE COUNT 37986 a 57251 c 54334 g 38738 t 665 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.25e+04 Length: 189014
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0
US-10-008-355-25 (1-10) x AC051649 (1-189014)
QY 1 ThrGlyAsnSerGlySerPro 8
|||||

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 55890: contig of 55890 bp in length
* 55991 55990: gap of 100 bp
* 55991 58431: contig of 2441 bp in length
* 58432 58531: gap of 100 bp
* 58532 75875: contig of 17344 bp in length
* 75876 75975: gap of 100 bp
* 75976 80404: contig of 4429 bp in length
* 80405 80504: gap of 100 bp
* 80505 164904: contig of 84400 bp in length
* 164905 165004: gap of 100 bp
* 165005 189014: contig of 24010 bp in length.
FEATURES
Location/Qualifiers
source
1. 189014
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone="RP11-534I22"
/clone_lib="RP11-534I22"
Human Male BAC"
BASE COUNT 37986 a 57251 c 54334 g 38738 t 665 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.25e+04 Length: 189014
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0
US-10-008-355-25 (1-10) x AC051649 (1-189014)
QY 1 ThrGlyAsnSerGlySerPro 8
|||||

```

Db 120252 ACCGGGGCAACTCGGTTCCCT 120275

RESULT 39  
AC121339

LOCUS  
DEFINITION Homo sapiens clone RP11-454C12, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 4  
unordered pieces.

ACCESSION AC121339 AC021983  
VERSION AC121339.7 GI:21306526  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T.,  
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Farraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H.,  
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Saverly, G.,  
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 193946)  
Worley, K.C.

Direct Submission  
Submitted (17-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 193946)  
Worley, K.C.

Direct Submission  
Submitted (02-JUN-2002) Human Genome Sequencing Center, Department,  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Jun 1, 2002 this sequence version replaced gi:21281305.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HEDB  
Center clone name: RP11-454C12  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 191935 bases at least Q40  
Consensus quality: 192192 bases at least Q30  
Consensus quality: 192297 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 16167: contig of 16167 bp in length  
\* 16168 16267: gap of unknown length  
\* 16268 40248: contig of 23981 bp in length  
\* 40249 40348: gap of unknown length  
\* 40349 118053: contig of 77705 bp in length  
\* 118054 118153: gap of unknown length  
\* 118154 193946: contig of 75793 bp in length.  
-----  
FEATURES  
Location/Qualifiers  
1..193946  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-454C12"

BASE COUNT 59591 a 37028 c 37958 g 59014 t 355 others  
ORIGIN

Alignment Scores:  
Pred. No.: 2,29e+04 Length: 193946  
Score: 44.00 Matches: 7  
Percent Similarity: 90.00% Conservatives: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 81.48% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AC121339 (1-193946)

Qy 1 ThrGlyGlyAsnSerGlySerProValphe 10  
|||||  
Db 144923 ACTGGGGCAACTCAGCAAGTCCTATTAT 144952

RESULT 40  
AL845498 206924 bp DNA linear HTG 20-AUG-2002  
LOCUS Mus musculus chromosome 2 clone RP23-222D20, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 13 unordered pieces.  
ACCESSION AL845498  
VERSION AL845498.2 GI:22416285  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
Plumb, B.  
1 (bases 1 to 206924)  
Direct Submission  
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Aug 21, 2002 this sequence version replaced gi:22316228.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute



Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: BM222020  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 202978 bases at least Q40  
 Consensus quality: 203961 bases at least Q30  
 Consensus quality: 204810 bases at least Q20  
 Insert size: 205724; sum-of-contigs  
 Insert size: 210104; 2.1% error; agarose-fp  
 Quality coverage: 6.43x in Q20 bases; sum-of-contigs Quality  
 coverage: 6.36x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 \* 4791 4890: contig of 4790 bp in length  
 \* 4891 36862: contig of 31972 bp in length  
 \* 36863 36962: gap of 100 bp  
 \* 36963 47151: contig of 10189 bp in length  
 \* 47152 47251: gap of 100 bp  
 \* 47252 92317: contig of 45066 bp in length  
 \* 92318 92417: gap of 100 bp  
 \* 92418 98370: contig of 5953 bp in length  
 \* 98371 98470: gap of 100 bp  
 \* 98471 112296: contig of 13826 bp in length  
 \* 112297 112396: gap of 100 bp  
 \* 112397 134994: contig of 22598 bp in length  
 \* 134995 135094: gap of 100 bp  
 \* 135095 164037: contig of 28943 bp in length  
 \* 164038 164137: gap of 100 bp  
 \* 164138 170564: contig of 6427 bp in length  
 \* 170565 170664: gap of 100 bp  
 \* 170665 176222: contig of 5536 bp in length  
 \* 176223 176322: gap of 100 bp  
 \* 176323 190422: contig of 14100 bp in length  
 \* 190423 190522: gap of 100 bp  
 \* 190523 200515: contig of 9993 bp in length  
 \* 200516 200615: gap of 100 bp  
 \* 200616 206924: contig of 6309 bp in length.

## FEATURES

## source

1. .206924  
 /location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-222D20"  
 /clone\_lib="RPC1-23"

## misc\_feature

1. .4790  
 /note="assembly\_fragment:01840  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left"  
 4891. .36862  
 /note="assembly\_fragment:00633  
 fragment\_chain:1"  
 36963. 47151  
 /note="assembly\_fragment:02096  
 fragment\_chain:1"  
 47252. .92317  
 /note="assembly\_fragment:01642  
 fragment\_chain:2"  
 92418. .98370  
 /note="assembly\_fragment:01195  
 fragment\_chain:2"

misc\_feature 98471. .112296  
 /note="assembly\_fragment:00640  
 fragment\_chain:2"  
 112397. .134994  
 /note="assembly\_fragment:01501  
 fragment\_chain:2"  
 135095. .164037  
 /note="assembly\_fragment:01368  
 fragment\_chain:2"  
 164138. .170564  
 /note="assembly\_fragment:02050  
 fragment\_chain:2"  
 170665. .176222  
 /note="assembly\_fragment:00634  
 fragment\_chain:2"  
 176323. .190422  
 /note="assembly\_fragment:01617  
 fragment\_chain:2"  
 190523. .200515  
 /note="assembly\_fragment:00522  
 fragment\_chain:2"  
 200616. .206924  
 /note="assembly\_fragment:01313  
 fragment\_chain:2  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 57066 a 40977 c 44179 g 63500 t 1202 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.4e+04 Length: 206924  
 Score: 44.00 Matches: 8  
 Percent Similarity: 80.00% Conservative: 0  
 Best Local Similarity: 80.00% Mismatches: 2  
 Query Match: 81.48% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AL845498 (1-206924)

QY 1 ThrGlyGlyAsnSerGlySerProValphe 10

Db 145655 ACGGAGGTCAGTCGGGTGAGCCAGTGTTC 145684

RESULT 41

AC102236

LOCUS

AC102236 222941 bp DNA linear HTG 21-AUG-2002  
 Mus musculus clone RP24-100013, WORKING DRAFT SEQUENCE, 28  
 unordered pieces.

ACCESSION

AC102236

VERSION AC102236.2 GI:22381151

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE

house mouse.

Mus musculus

ORGANISM

REFERENCE

1 (bases 1 to 222941)

AUTHORS

Birren,B., Linton,L.,

Anderson,S., Barna,N.,

Brown,A., Camarata,J.,

Choepel,Y., Colangelo,M.,

Cooke,P., DeArelano,K.,

Ferreira,P., FitzHugh,W.,

Ginde,S., Gord,S.,

Hagdo,B., Heaford,A.,

Jones,C., Kamat,A.,

Lamazares,R., Landers,T.,

MacLean,C., Macdonald,P.,

McCarthy,M., McEwan,P.,

Meneus,L., Mihova,T.,

Murphy,T., Naylor,J.,

Nguyen,C.,

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 222941)  
 Birren,B., Nusbbaum,C. and Lander,E.  
 Mus musculus, clone RP24-100013  
 Unpublished

2 (bases 1 to 222941)  
 Birren,B., Linton,L., Nusbbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagdo,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Kamat,A., Karatas,A., Kells,C., Labocque,K.,  
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,  
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE

JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 222941)  
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE

JOURNAL Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Aug 21, 2002 this sequence version replaced gi:17061322.  
 All repeats were identified using RepeatMasker:  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L18248

Center clone name: 100\_Q\_13

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 215030 bases at least Q40  
 Consensus quality: 218043 bases at least Q30  
 Consensus quality: 219294 bases at least Q20  
 Insert size: 210000; agarose-fp  
 Insert coverage: 220241; sum-of-contigs  
 Quality coverage: 8.1 in Q20 bases; agarose-fp  
 Quality coverage: 7.8 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence \* as soon as it is available and the accession number will be preserved.

\* 1 165: contig of 165 bp in length  
 \* 166 265: gap of 100 bp  
 \* 266 469: contig of 204 bp in length  
 \* 470 569: gap of 100 bp  
 \* 570 585: contig of 16 bp in length  
 \* 586 685: gap of 100 bp  
 \* 686 816: contig of 131 bp in length

\* 817 916: gap of 100 bp  
 \* 917 1581: contig of 665 bp in length  
 \* 1582 1681: gap of 100 bp  
 \* 1682 2308: contig of 627 bp in length  
 \* 2309 2408: gap of 100 bp  
 \* 2409 3613: contig of 1205 bp in length  
 \* 3614 3713: gap of 100 bp  
 \* 3714 4744: contig of 1031 bp in length  
 \* 4745 4844: gap of 100 bp  
 \* 4845 5791: contig of 947 bp in length  
 \* 5792 5891: gap of 100 bp  
 \* 5892 6751: contig of 860 bp in length  
 \* 6752 6851: gap of 100 bp  
 \* 6852 8046: contig of 1195 bp in length  
 \* 8047 8146: gap of 100 bp  
 \* 8147 9287: contig of 1151 bp in length  
 \* 9288 9397: gap of 100 bp  
 \* 9398 10587: contig of 1190 bp in length  
 \* 10588 10687: gap of 100 bp  
 \* 10689 12073: contig of 1386 bp in length  
 \* 12074 12173: gap of 100 bp  
 \* 12174 13315: contig of 1142 bp in length  
 \* 13316 13415: gap of 100 bp  
 \* 13416 15361: contig of 1946 bp in length  
 \* 15362 15461: gap of 100 bp  
 \* 15462 17033: contig of 1572 bp in length  
 \* 17034 17133: gap of 100 bp  
 \* 17134 19542: contig of 2409 bp in length  
 \* 19543 19642: gap of 100 bp  
 \* 19643 21733: contig of 2111 bp in length  
 \* 21734 21853: gap of 100 bp  
 \* 21854 23041: contig of 1188 bp in length  
 \* 23042 23141: gap of 100 bp  
 \* 23142 27602: contig of 4461 bp in length  
 \* 27603 27702: gap of 100 bp  
 \* 27703 31693: contig of 3991 bp in length  
 \* 31694 31793: gap of 100 bp  
 \* 31794 39800: contig of 8007 bp in length  
 \* 39801 39900: gap of 100 bp  
 \* 39901 54439: contig of 14539 bp in length  
 \* 54440 54539: gap of 100 bp  
 \* 54540 80457: contig of 25918 bp in length  
 \* 80458 80557: gap of 100 bp  
 \* 80558 114212: contig of 33655 bp in length  
 \* 114213 114312: gap of 100 bp  
 \* 114313 163570: contig of 49258 bp in length  
 \* 163571 163670: gap of 100 bp  
 \* 163671 222941: contig of 59271 bp in length.

## FEATURES

Location/Qualifiers  
 1. .222941  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP24-100013"  
 /clone\_lib="RPCr-24 Male Mouse BAC"  
 1. .165  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 misc\_feature 266. 469  
 /note="assembly\_fragment"  
 misc\_feature 570. 585  
 /note="assembly\_fragment"  
 misc\_feature 686. 816  
 /note="assembly\_fragment"  
 misc\_feature 917. 1581  
 /note="assembly\_fragment"  
 misc\_feature 1682. 2308  
 /note="assembly\_fragment"  
 misc\_feature 2409. 3613  
 /note="assembly\_fragment"  
 misc\_feature 3714. 4744  
 /note="assembly\_fragment"  
 misc\_feature 4845. 5791

```

/note="assembly_fragment"
5892..6751
/note="assembly_fragment"
6852..8046
/note="assembly_fragment"
8147..9297
/note="assembly_fragment"
9398..10587
/note="assembly_fragment"
10688..12073
/note="assembly_fragment"
12174..13315
/note="assembly_fragment"
13416..15361
/note="assembly_fragment"
15462..17033
/note="assembly_fragment"
17134..19542
/note="assembly_fragment"
19643..21753
/note="assembly_fragment"
21854..23041
/note="assembly_fragment"
23142..27602
/note="assembly_fragment"
27703..31693

```

```

Alignment Scores:
Pred. No.: 2,54e+04 Length: 222941
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 2 Gaps: 0

```

US-10-008-355-25 (1-10) x AC102236 (1-222941)

QY 2 GlyGlyAsnSerGlySerProValPhe 10

Db 199722 GGGGGAGATTCTGGAAGTCCTGTAATT 199748

RESULT 42  
AE003695/c

LOCUS AE003695 225655 bp DNA linear INV 05-OCT-2000  
DEFINITION Drosophila melanogaster genomic scaffold 14200001386035 section 20  
of 105, complete sequence.

ACCESSION AE003695 AE002708

VERSION AE003695.1 GI:7299572

KEYWORDS HTG.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 225655)

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,  
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,  
Brandon,R.C., Rogers,Y.H., Blazer,J.R.G., Champen,M., Pfeiffer,B.D.,  
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,  
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,  
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,  
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,  
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,  
Fleischmann,W., Fosier,C., Gabrielian,A.E., Garg,N.S.,  
Gelbart,W.N., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,

```

Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Liu,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,C.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nuskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kimnos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,O.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
10731132
2 (bases 1 to 225655)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
FEATURES
source
location/Qualifiers
1..225655
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
complement(<4520..>6506)
/gene="CG4830"
/map="87B5-87B6"
/db_xref="FLYBASE:FBan0004830"
/db_xref="FLYBASE:FBgn0037996"
complement(join(4520..5471,5561..5677,5747..5933,
6046..6253,6325..>6506))
/gene="CG4830"
/product="CT15487"
/db_xref="FLYBASE:FBan0004830"
/db_xref="FLYBASE:FBgn0037996"
complement(join(4520..5471,5561..5677,5747..5933,
6046..6253,6325..>6506))
/gene="CG4830"
/note="CG4830 gene product"
/codon_start=1
/protein_id="AAF54758.1"
/db_xref="GI:7299573"
/db_xref="FLYBASE:FBan0004830"
/db_xref="FLYBASE:FBgn0037996"
/translation="MSKLPQSYDAENKWKIPQNNPKPETSLSGLRIIFKNNRPNPKN
VQIQDSGEVEVTFQALPTWAIQAKSLGHDHDIIGISARNITYTPTAVACFF
HCTPQSDPALEESTLKLHYNISKPIFTDADHYDKLYSATSAFPIILTTGTQD
GVLSIQDLAPTKTEFFQPTPLKEGSPOTVAILTSSGTTGMPKAVCISNDILTOFTV
FVNGVDTFISASLDWITGLWATIEFVNGCTRIISSKPFADYFVLYSKYSIYAL
IPEHCHSLDCLPTAPPEKGLSLTKLNFQGGRTQATVERVKRLAPNGVNSSYGME
EATKAMQDEGWFHTGDGMYGFDEDDYLYMTDRKKEYLKWKLQMPAEVAYIDELPE
VKRCVIGDEFTQDGPALGVLRREDNLTAAQQVIDHVAKRLPDITQKLRAGVQFAD
EIPQNHGKVVRRYARDLFFVALSKKN"
<10971..>11742
/gene="CG14742"
/evidence=not_experimental
/map="87B5-87B6"
/db_xref="FLYBASE:FBan0014742"
/db_xref="FLYBASE:FBgn0037997"
join(<10971..11039,11587..>11742)
/gene="CG14742"
/product="CT34536"

```



```

----- Project Information -----
Center project name: H319
Center clone name: SCb-212e3
----- Summary Statistics -----
Sequencing vector: M13; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 234404 bases at least Q40
Consensus quality: 234413 bases at least Q30
Consensus quality: 234428 bases at least Q20
Quality coverage: 5.81 x in Q20 bases; sum-of-contrigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES             Location/Qualifiers
     source            1. 234431
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="8"
                        /map="8p12"
                        /clone="SCb-212e3"
BASE COUNT    62993 a 56257 c 55890 g 59291 t
ORIGIN
Alignment Scores:
Pred. No.:          2.63e+04      Length:      234431
Score:              44.00         Matches:      8
Percent Similarity: 100.00%       Conservative: 0
Best Local Similarity: 100.00%     Mismatches:    0
Query Match:        81.48%         Indels:        0
DB:                  9            Gaps:          0

US-10-008-355-25 (1-10) x AF311103 (1-234431)

Oy 1 ThrGlyGlyAsnSerGlySerPro 8
Db 229028 ACAGAGGGGACAGTGGCAGCCCT 229051

RESULT 44
AC126675
LOCUS       AC126675             279166 bp    DNA     linear     HTG 08-JUL-2002
DEFINITION Mus musculus chromosome UNK clone RP23-430P11, WORKING DRAFT
SEQUENCE   56 unordered pieces.
AC126675
VERSION     AC126675.1 GI:21703664
KEYWORDS    HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            McPherson,J.D. and Waterston,R.H.
            1 (bases 1 to 279166)
            The sequence of Mus musculus clone
            Unpublished
            JOURNAL
            2 (bases 1 to 279166)
            McPherson,J.D. and Waterston,R.H.
            Direct Submission
            AUTHORS
            Submitted (08-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
            JOURNAL
            COMMENT
            ----- Genome Center -----
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BAO430P11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 262726 bases at least Q40
Consensus quality: 269647 bases at least Q30
Consensus quality: 275322 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 273932; sum-of-contrigs
Quality coverage: 0.00 in Q20 bases; agarose-fp
Quality coverage: 4.57 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1262: contrig of 1262 bp in length
* 1263 1362: gap of unknown length
* 1363 2479: contrig of 1117 bp in length
* 2480 2579: gap of unknown length
* 2580 3655: contrig of 1076 bp in length
* 3656 3755: gap of unknown length
* 3756 5131: contrig of 1376 bp in length
* 5132 5231: gap of unknown length
* 5232 6493: contrig of 1261 bp in length
* 6493 6592: gap of unknown length
* 6593 7734: contrig of 1142 bp in length
* 7735 9435: contrig of 1601 bp in length
* 9436 9535: gap of unknown length
* 9536 10753: contrig of 1218 bp in length
* 10754 10853: gap of unknown length
* 10854 11874: contrig of 1021 bp in length
* 11875 11974: gap of unknown length
* 11975 13469: contrig of 1495 bp in length
* 13470 13569: gap of unknown length
* 13570 14673: contrig of 1104 bp in length
* 14674 14773: gap of unknown length
* 14774 16482: contrig of 1709 bp in length
* 16483 16582: gap of unknown length
* 16583 17774: contrig of 1192 bp in length
* 17775 17874: gap of unknown length
* 17875 18984: contrig of 1110 bp in length
* 18985 19084: gap of unknown length
* 19085 20671: contrig of 1587 bp in length
* 20672 20771: gap of unknown length
* 20772 22144: contrig of 1373 bp in length
* 22145 22244: gap of unknown length
* 22245 23825: contrig of 1581 bp in length
* 23826 23925: gap of unknown length
* 23926 25361: contrig of 1436 bp in length
* 25362 25461: gap of unknown length
* 25462 27152: contrig of 1691 bp in length
* 27153 27252: gap of unknown length
* 27253 29103: contrig of 1851 bp in length
* 29104 29203: gap of unknown length
* 29204 30595: contrig of 1392 bp in length
* 30596 30696: gap of unknown length
* 30697 32326: contrig of 1631 bp in length
* 32327 32426: gap of unknown length
* 32427 34498: contrig of 2072 bp in length
* 34499 34598: gap of unknown length
* 34599 36514: contrig of 1916 bp in length
* 36515 36614: gap of unknown length

```

```

* 36615 38037: contig of 1423 bp in length
* 38038 38137: gap of unknown length
* 38138 40447: contig of 2310 bp in length
* 40448 40547: gap of unknown length
* 40548 42501: contig of 1954 bp in length
* 42502 42601: gap of unknown length
* 42602 43933: contig of 1332 bp in length
* 43934 44033: gap of unknown length
* 44034 46164: contig of 2131 bp in length
* 46165 46264: gap of unknown length
* 46265 47907: contig of 1643 bp in length
* 47908 48007: gap of unknown length
* 48008 49910: contig of 1903 bp in length
* 49911 50010: gap of unknown length
* 50011 52781: contig of 2771 bp in length
* 52782 52881: gap of unknown length
* 52882 55340: contig of 2459 bp in length
* 55341 55440: gap of unknown length
* 55441 58838: contig of 3398 bp in length
* 58839 58938: gap of unknown length
* 58939 61431: contig of 2493 bp in length
* 61432 61531: gap of unknown length
* 61532 64127: contig of 2596 bp in length
* 64128 64227: gap of unknown length
* 64228 67688: contig of 3461 bp in length
* 67689 67788: gap of unknown length
* 67789 71092: contig of 3304 bp in length
* 71093 71192: gap of unknown length
* 71193 74415: contig of 3223 bp in length
* 74416 74515: gap of unknown length
* 74516 77637: contig of 3122 bp in length
* 77638 77737: gap of unknown length
* 77738 81100: contig of 3363 bp in length
* 81101 81200: gap of unknown length
* 81201 84740: contig of 3540 bp in length
* 84741 84840: gap of unknown length
* 84841 88327: contig of 3487 bp in length
* 88328 88427: gap of unknown length
* 88428 90210: contig of 1783 bp in length
* 90211 90310: gap of unknown length
* 90311 93507: contig of 3197 bp in length
* 93508 93607: gap of unknown length
* 93608 97048: contig of 3441 bp in length
* 97049 97148: gap of unknown length
* 97149 101273: contig of 4125 bp in length
* 101274 101373: gap of unknown length
* 101374 108597: contig of 7224 bp in length
* 108598 108697: gap of unknown length
* 108698 117560: contig of 8863 bp in length
* 117561 117660: gap of unknown length
* 117661 125391: contig of 7731 bp in length
* 125392 125491: gap of unknown length
* 125492 140929: contig of 15438 bp in length
* 140930 141029: gap of unknown length
* 141030 154376: contig of 13347 bp in length
* 154377 154476: gap of unknown length
* 154477 169499: contig of 15023 bp in length
* 169500 169599: gap of unknown length
* 169600 184510: contig of 14911 bp in length
* 184511 184610: gap of unknown length
* 184611 208772: contig of 24162 bp in length
* 208773 208872: gap of unknown length
* 208873 279166: contig of 70294 bp in length.

FEATURES
    source
    1..279166
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /chromosome="UNK"
        /clone="RP23-430P11"
        1..1262
        /note="assembly_name:Contig5"
        1363..2479
        /note="assembly_name:Contig21"

```

```

misc_feature 2580..3655
    /note="assembly_name:Contig25"
misc_feature 3756..5131
    /note="assembly_name:Contig32"
misc_feature 5232..6492
    /note="assembly_name:Contig35"
misc_feature 6593..7734
    /note="assembly_name:Contig38"
misc_feature 7835..9435
    /note="assembly_name:Contig39"
misc_feature 9536..10753
    /note="assembly_name:Contig44"
misc_feature 10854..11874
    /note="assembly_name:Contig48"
misc_feature 11975..13469
    /note="assembly_name:Contig51"
misc_feature 13570..14673
    /note="assembly_name:Contig52"
misc_feature 14774..16482
    /note="assembly_name:Contig57"
misc_feature 16583..17774
    /note="assembly_name:Contig58"
misc_feature 17875..18984
    /note="assembly_name:Contig59"
misc_feature 19085..20671
    /note="assembly_name:Contig61"
misc_feature 20772..22144
    /note="assembly_name:Contig62"
misc_feature 22245..23825

Alignment Scores:
Pred. No.: 2.99e+04
Score: 44.00
Percent Similarity: 100.00%
Best Local Similarity: 88.89%
Query Match: 81.48%
DB: 2

US-10-008-355-25 (1-10) x AC126675 (1-279166)
QY 2 GlyGlyAsnSerGlySerProValphe_10
Db 163645 GGTGGGCTTCTGGATCCCCAGTATTT 163671

RESULT 45
AP002995/c AP002995 346897 bp DNA linear BCT 15-MAY-2001
LOCUS Mesorhizobium loti DNA, complete genome, section 2/21.
DEFINITION Mesorhizobium loti DNA, complete genome, section 2/21.
ACCESSION AF002995 BA000012
VERSION AP002995.2 GI:14021442
KEYWORDS
SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.
ORGANISM Mesorhizobium loti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
    1 (sites)
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
    Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
    Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
    Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
    Takeuchi,C., Yamada,M. and Tabata,S.
    Complete genome structure of the nitrogen-fixing symbiotic
    bacterium Mesorhizobium loti
    JOURNAL Mesorhizobium loti
    DNA Res. 7 (6), 331-338 (2000)
    MEDLINE 21082930
    REFERENCE 2 (bases 1 to 346897)
    Direct Submission
    Kaneko,T.
    Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
    Institute, The First Laboratory for Plant Gene Research; Yana
    1532-3, Kisarazu, Chiba 292-0812, Japan
    (E-mail:kaneko@kazusa.or.jp,
    URL:http://www.kazusa.or.jp/rhizobase/,

```

Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)  
 On May 11, 2001 this sequence version replaced gi:11994963.  
 COMMENT Location/Qualifiers  
 FEATURES source  
 1..346897  
 /organism="Mesorhizobium loti"  
 /strain="MAFF303099"  
 /db\_xref="taxon:381"  
 74..160  
 /note="trns-TGA"  
 complement(240..1109)  
 /gene="msl8587"  
 complement(240..1109)  
 /gene="msl8587"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48055.1"  
 /db\_xref="GI:14021443"  
 /translation="MYDQTTTGITALPYVVRIRAGDVANQAVHKGTGFYHFRREG  
 AGNPLIVSNKHVLVKGLEFDFALLDNOGRVFGPANTFRLTKNQLPIFPHDPPTV  
 DLAAPIQPIVEMRGQKEIFNLLGRETIAPDNIIQAISLVAIVMGFPGSIMDE  
 KNNLPVRRGIILAVNLTDRGETNFVVDIAAAGSSGSPVFAFFENMLPDGSGGVTM  
 LAQPOVFLGVLSGSPMTATGKIIVPVPTSDQIAQTOMLHLGLVCYKASRVEELVE  
 TIRAKISEPTASS"  
 complement(1118..1657)  
 /gene="ml10476"  
 complement(1118..1657)  
 /gene="ml10476"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48056.1"  
 /db\_xref="GI:14021444"  
 /translation="MFPISALIGFEKGOWGALALAGAMAQTAEDNFRIEWDREE  
 KNLVETICRSPDSFTSOAAWQAIIRRPCGMVLIHNSRVHMEKIITPGPKVPOTIV  
 DCSVAGQDVALGDLREWHKLRANCKNSHAEVPAALIKYGEALFSTVERALFC  
 TCGDRGGVRLRIHKLPRN"  
 2193..2558  
 /gene="mlr0478"  
 2193..2558  
 /gene="mlr0478"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48057.1"  
 /db\_xref="GI:14021445"  
 /translation="MALALGGALMKWRLYFVVMTLAVTSGCAADYLNNDYDTVLASGD  
 ANHTNSLQTVDPFPNPNNSNTHIEGQRIQVGVQYRSPISAGGYSNCPTDNSVA  
 ADGSHGGRKRSRPGGAM"  
 2651..3412  
 /gene="mlr0479"  
 2651..3412  
 /gene="mlr0479"  
 /codon\_start=1  
 /transl\_table=11  
 /product="succinoglycan biosynthesis protein; ExoI"  
 /protein\_id="BAB48058.1"  
 /db\_xref="GI:14021446"  
 /translation="MPLRLVLTVAASALATQFVMHNGSSILPEINLNLKPKPVAI  
 SGVASVLDGDTTEHGQVRVFNIGIDAPESAQYCDKAKGREYPCGRSSAALLGFLAAS  
 KPQCAFTWDRIGREVGCTRADGSSVAAMVHEHQALDWPRIYNSGAYAAQAAKKA  
 AKVGLWIGSFQAPWQRAQSDDAQTPTAPLATGTGNGPNCNIKGNISAGERIYHLP  
 GKQYXAVTVISQNGKRWFCSEAEAAVAGWRRSKR"  
 3429..3782  
 /gene="mlr0480"  
 3429..3782  
 /gene="mlr0480"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48059.1"  
 /db\_xref="GI:14021447"  
 /translation="MKPHRAVAATAALSHSSGRNVSVYSYSELKYVNIDASVQGNV  
 VDADYNNNSCHDGTLPSPSYHYHGQNSHIDLQPSQGGQYDGYDYGNSNHPFELTVSGNSA  
 ELYDCEGFFSFT"  
 complement(5516..5947)  
 /gene="ml10481"  
 complement(5516..5947)  
 /gene="ml10481"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48060.1"  
 /db\_xref="GI:14021448"  
 /translation="MDNSDRGTGKDPSTGRFLPGHSFKVKVGSRSRVSRAIDTANEHA  
 ATAMTLMVGKMLAGDIRAAEILIRYGADPFLPKNRTVELPVDENPTSENLAVALSAA  
 WSGEISPDPAKQLLDIAAAGKGTIEDAELLRRVEQLEAVKR"  
 complement(6065..6883)  
 /gene="ml10482"  
 complement(6065..6883)  
 /gene="ml10482"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48061.1"  
 /db\_xref="GI:14021449"  
 /translation="MIKDFKRIDGASSRGSLSFQWNYNGSVISAFWNEYSTMSPDIDA  
 AASALNGVLAAGNAVIGAEARKRTDEIRDSLEAIARIYVAAPFRLDRALDERNS  
 VGISRNSLAQPTIPIPTPTSTAADSLEQERVRMLDGFQDGEAFDWMQSPGIVL  
 NAVAPSIDLTFISKDQQLAGLIVEFAIRKWARTIDPNSDNILATAPNLKKNMSDDL  
 VRDLAKARLDGLKREQAVSDAEKLLAVATFASQAGNLTPDEAFKMLMGQKVA"  
 complement(7133..7753)  
 /gene="ml10483"  
 complement(7133..7753)  
 /gene="ml10483"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48062.1"  
 /db\_xref="GI:14021450"  
 /translation="MANYVILRLVMKNRASTITADHANFAKIQGTADHAFIWLSGRT  
 REWASPNPSEFYRDECDMDVSYEGNRTTFSPDKTYQSYSPMFDAAATTDGHPFVW  
 WGTLLGCADNMVSVSTHFSFGIINGVMSASVDQSDNNQFEDSDSTVYSEMILDRHLQPW  
 RDPPTGVARQSKLEPAILRALVKLYRSKDTERRWTAPNQFAENIH"  
 complement(7827..10169)  
 /gene="ml10485"  
 complement(7827..10169)  
 /gene="ml10485"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48063.1"  
 /db\_xref="GI:14021451"  
 /translation="MRDGSSVRLPELLPYRQFVWVRLEOKGPKPKAMPYSANGGR  
 ADITNPATWSYPTAVRALSGQYAGIGFVFAENDPFFVIDLGDGCRDATTGELSAEAA  
 ALTRMPNNAWELSSQSEGIHIFGRCDKVALGGRKNAWAGCEFYRTKRFVAPGPGI  
 QGNIDIDWTNOLAILVPRGVGEODASQEAVERDPRIWIGDDDEELLRLRFLASTGTTK  
 AKLQASPTNLQMEHAALAAQYFPSATUKPDPHSSADQALANLAFWTGRTGYRIIR  
 LFGRSALGRKRTYRAAYNSVTNAAVRSQCTVYGGNGOERSQVQENRIGSDIG  
 ADITHVTMLDEMLQDLWFIHRKSGGCAIADSTLNLVLPATAQRSYASRYEAAST  
 IDPKTDGFKRYLPVFNOWFQSPSRKSDVITWKPNGAVICEAPESQAGFNWRFQ  
 PYELPDGOSLVEVWEHLAYLPIEAEARVFEQWLAHIVORPEELPOSAYLFTRET  
 GIGRWLSSVLVRLRGYVAAGIDIOEMLDGTGNGRLSRKLLAVDEAKAGMGQYGRW  
 SHSEKLDHFMEMRRINRKHGLEFVEWNCRRWLFFSNNDLPPFASDRRRNVNAMP  
 THRSQDPYERYIYQIANPDFTSAVRKRLLELDISGFKPGCAIPDMNEARKKLGAMAS  
 LDNLNLAAFRDEWPGPLAERSDVINYLHVSIAPKLPEKTIINKLIEQAGMKLVERISIG  
 PNKYRFVYVVDYTAQOIEITNPWAYAEAVNARKGYGFRQG"  
 complement(10162..10392)  
 /gene="ml10486"  
 complement(10162..10392)  
 /gene="ml10486"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48064.1"  
 /db\_xref="GI:14021452"

/translation="MECVAMAGNLRVREMNOQSAIVLSIPQVCATLGISRPTLDRIYKID  
 NPSFPRKKIGPRLVGLRSDIETIYIGSLPDA"  
 complement(10595..11803)  
 /gene="mli10487"  
 complement(10595..11803)  
 /gene="mli10487"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable prophage integrase"  
 /protein\_id="BAB48065.1"  
 /db\_xref="GI:114021453"  
 /translation="MITIREIQSLVHKGPGQMSHTKALYFFFIAPSGRPFWAIRYMLK  
 GRRQSLSDISIPATVRELKQIELLSETEREKVKGDIDPLAEREAKAASPEVROLV  
 TFKSVADLYTAANRGGWKNKAKHVQOQENLATYVYPVIGHLEPOSITTAELVKILOPL  
 WAKRTASVDRSPKRTVETIGAARAKGMSEKPEIWAHNHNPARNADNLEFWLNGKQKQK  
 HFSAKMPYETPAQAOSTAREDYSTKALOPILITATPTNETLGAENSEIDTEKAVETI  
 PASMKSGREHVRPDLSEALKLVASLLIREGNYLPFGAIRNKSLSNMAALLMVRGLK  
 LGHYVPHGFSRPAEDWAAEETVYPNHIVEMALAHTIRNKVVEAAYRRGDLMKKRRRLME  
 DWSAYLASAR"  
 12686..13279  
 /gene="mli0488"  
 12686..13279  
 /gene="mli0488"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48066.1"

Alignment Scores:  
 Pred. No.: 3,51e+04 Length: 346897  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x AP002995 (1-346897)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10  
 Db 530 GGTGATCCTCAGATCGCCGGTTT 504  
 RESULT 46  
 AX431912  
 LOCUS AX431912 534 bp DNA linear PAT 28-JUN-2002  
 DEFINITION Sequence 327 from Patent WO0229113.  
 ACCESSION AX431912  
 VERSION AX431912.1 GI:21656716  
 KEYWORDS  
 SOURCE Bacillus licheniformis.  
 ORGANISM Bacillus licheniformis  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 REFERENCE 1  
 AUTHORS Berka, R. and Clausen, I. G.  
 TITLE Methods for monitoring multiple gene expression  
 JOURNAL Patent: WO 0229113-A 327 11-APR-2002;  
 Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)  
 FEATURES  
 source  
 1..534  
 /organism="Bacillus licheniformis"  
 /db\_xref="taxon:1402"  
 BASE COUNT 158 a 144 c 128 g 104 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 436 Length: 534  
 Score: 43.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 79.63% Indels: 0  
 DB: 6 Gaps: 0

US-10-008-355-25 (1-10) x AX431912 (1-534)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10  
 Db 357 GGAGGACAAACGGGTTCCACCGGTATTC 383  
 RESULT 47  
 BACGASP  
 LOCUS BACGASP 1448 bp DNA linear BCT 01-FEB-2000  
 DEFINITION B.licheniformis glutamic acid-specific protease gene.  
 ACCESSION D10060  
 VERSION D10060.1 GI:216263  
 KEYWORDS glutamic acid-specific protease.  
 SOURCE Bacillus licheniformis (strain:ATCC14580) DNA.  
 ORGANISM  
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 REFERENCE 1 (bases 1 to 1448)  
 AUTHORS Kakudo, S., Kikuchi, N., Kitadokoro, K., Fujiwara, T., Nakamura, E.,  
 Okamoto, H., Shin, M., Tamaki, M., Teraoka, H., Tsuzuki, H. and  
 Yoshida, N.  
 TITLE Purification, characterization, cloning, and expression of a  
 glutamic acid-specific protease from Bacillus licheniformis ATCC  
 14580  
 J. Biol. Chem. 267 (33), 23782-23788 (1992)  
 MEDLINE 93054737  
 REFERENCE 2 (bases 1 to 1448)  
 AUTHORS Nakamura, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-DEC-1991) Etsuo Nakamura, Shionogi & Co., Ltd.,  
 Shionogi Research Laboratories; 12-4 Sagisu, 5-chome, Fukushima-ku,  
 Osaka, Osaka 553, Japan (E-mail:ishizaki@shlogw.lab.shionogi.co.jp,  
 Tel:06-458-5861(ex.571), Fax:06-458-0987)  
 COMMENT Submitted (09-DEC-1991) to DDBJ by:  
 Etsuo Nakamura  
 Shionogi Research Laboratories  
 Ltd.  
 Fukushima-ku, Osaka 553  
 Japan  
 Phone: 06-458-5861 x571  
 Fax: 06-458-0987  
 E mail: nakamura@shlogw.lab.shionogi.co.jp@ddlabs.co.jp.  
 FEATURES  
 source  
 1..1448  
 /organism="Bacillus licheniformis"  
 /strain="ATCC14580"  
 /db\_xref="taxon:1402"  
 323..1273  
 /gene="Blase"  
 323..1273  
 /gene="Blase"  
 /note="ttg start codon, prepro sequence"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glutamic acid-specific protease"  
 /protein\_id="BAA00949.1"  
 /db\_xref="GI:216264"  
 /translation="MVSKSKVRKGLITGLIGISYSLGMHPAQAPSPHTPVSSDPSY  
 KAETSVYDPEINIKSDQYGLYSKAFGTGKVNETHKAEKKSPAKAPYSIKSVIGSDDR  
 TRVNTTAYPRALVHVISSISCTGMIKPTKATAGCIYDTSSGSACTATVSPG  
 RGTSTPYGSVKSTRYRIFSGWRSGNTNDYGAIELSEPIGNTVGFYGYSTYTSVLVG  
 TTVTISGYPGDKTAGTQWQHSGPIAISEYIKLOYAMDITYGQSGSPVFEQSSSRNTCS  
 GPCSLAVHTNGVYGGSSYNRGTRITRITKEVFDNLTNMKNLSAQ"  
 605..1270  
 /gene="Blase"  
 /product="glutamic acid-specific protease mature peptide"  
 BASE COUNT 431 a 360 c 294 g 363 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 906 Length: 1448  
 Score: 43.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 79.63% Indels: 0



Unknown.  
Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 1448)  
AUTHORS Teraoka,H., Tanaki,M., Nakamura,E., Shin,M., Yoshida,N.,  
Tsuzuki,H., Fujiwara,T. and Matsumoto,K.  
TITLE Protease  
JOURNAL  
FEATURES Patent: US 5459064-A 1 17-OCT-1995;  
Location/Qualifiers  
1..1448  
/organism="unknown"  
BASE COUNT 431 a 360 c 294 g 363 t  
ORIGIN

Alignment Scores:  
Pred. No.: 906 Length: 1448  
Score: 43.00 Matches: 8  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 1  
Query Match: 79.63% Indels: 0  
DB: 6 Gaps: 0

US-10-008-355-25 (1-10) x 115101 (1-1448)

QY 2 GlyGlyAsnSerGlySerProValphe 10  
||||| |||||||||  
Db 1094 GGAGGACAAACGGCTTCACCGGTATTC 1120

RESULT 50  
AL5900093  
LOCUS  
DEFINITION Human DNA sequence from clone Rp11-240D10 on chromosome 1, complete  
sequence.  
ACCESSION AL5900093 AC027618  
VERSION AL5900093.10 G:20302181  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Almeida,J.  
Direct Submission  
Submitted (20-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
[humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk) Clone requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Apr 24, 2002 this sequence version replaced gi:16973058.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr1>  
Rp11-240D10 is from the library RPCR-11.1 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6

Draft Sequence Produced by Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
<http://genome.wustl.edu/gsc/index.shtml>.

FEATURES

source  
1. .62485  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="Rp11-24OD10"  
/clone\_lib="RPCI-11.1"

BASE COUNT 15588 a 14738 c 16165 g 15994 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.43e+04 Length: 62485  
Score: 43.00 Matches: 7  
Percent Similarity: 88.89% Conservative: 1  
Best Local Similarity: 77.78% Mismatches: 1  
Query Match: 79.63% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AL590093 (1-62485)

Qy 1 ThrGlyGlyAsnSerGlySerProVal 9  
|||||  
Db 45330 ACTGGAGGAATTCAGGAGAGCCCATC 45356

Search completed: May 23, 2003, 14:10:21  
Job time : 1631 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 13:44:18 ; Search time 83 Seconds  
(without alignments)  
159.092 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TGNNGSPVF 10

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-O/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110505\_3201/app\_query.fasta\_1.199  
-DB=PublishedApplications\_NA\_QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10008355\_QCGN\_1.1.99 @runat\_16052003\_110505\_3201  
-NCPU=6 -ICPU=3 -NO\_XLPHY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_NA:\*\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	1019	7	US-08-781-986A-752
2	54	100.0	2139	9	US-10-008-355-1
3	45	83.3	475	9	US-09-918-995-37867
4	44	81.5	423	9	US-09-764-891-18

Sequence 261, App	4951	81.5	44	10	US-09-070-927A-261
Sequence 327, App	534	79.6	43	10	US-09-974-300-327
Sequence 5, Appli	152	77.8	7	9	US-10-070-676-5
Sequence 31306, A	471	77.8	8	9	US-09-918-995-31306
Sequence 1705, Ap	450	74.1	40	10	US-09-867-701-1705
Sequence 13875, A	483	74.1	10	9	US-09-918-995-13875
Sequence 243, App	7441	74.1	10	10	US-10-079-854-243
Sequence 243, App	7441	74.1	10	10	US-09-764-878-243
Sequence 5572, Ap	235	72.2	39	13	US-09-960-352-5672
Sequence 22868, A	479	72.2	14	9	US-09-918-995-22868
Sequence 9954, Ap	584	72.2	39	15	US-09-864-761-9954
Sequence 136, App	1173	72.2	39	16	US-09-893-519A-136
Sequence 75, Appl	1775	72.2	39	17	US-10-001-887-75
Sequence 111, App	2109	72.2	39	18	US-09-764-868-111
Sequence 683, App	3255	72.2	39	19	US-10-025-380-683
Sequence 683, App	3255	72.2	39	20	US-09-922-217-683
Sequence 683, App	3255	72.2	39	21	US-09-833-263-683
Sequence 11, Appl	123526	72.2	39	22	US-09-910-185-11
Sequence 1415, Ap	218	70.4	38	10	US-09-815-343-1415
Sequence 676, App	222	70.4	38	24	US-09-815-343-676
Sequence 1285, Ap	222	70.4	38	25	US-09-815-343-1285
Sequence 614, App	243	70.4	38	26	US-09-815-343-614
Sequence 729, App	253	70.4	38	27	US-09-815-343-729
Sequence 5591, Ap	319	70.4	38	28	US-09-796-692-5591
Sequence 5591, Ap	319	70.4	38	29	US-10-040-862-5591
Sequence 6270, Ap	320	70.4	38	30	US-09-796-692-6270
Sequence 6270, Ap	320	70.4	38	31	US-10-040-862-6270
Sequence 591, App	454	70.4	38	32	US-09-770-444-591
Sequence 340, App	460	70.4	38	33	US-09-918-995-340
Sequence 5230, Ap	474	70.4	38	34	US-09-867-701-5230
Sequence 6501, Ap	491	70.4	38	35	US-09-783-590-6501
Sequence 4015, Ap	508	70.4	38	36	US-09-783-590-4015
Sequence 123, App	531	70.4	38	37	US-10-076-622-123
Sequence 123, App	531	70.4	38	38	US-09-604-287A-123
Sequence 123, App	531	70.4	38	39	US-08-339-338-123
Sequence 123, App	531	70.4	38	40	US-10-007-805-123
Sequence 27317, A	562	70.4	38	41	US-09-918-995-27317
Sequence 33, Appl	1171	70.4	38	10	US-09-808-701-14
Sequence 332, App	1514	70.4	38	10	US-10-043-487-33
Sequence 3293, A	2290	70.4	38	10	US-09-764-864-332
Sequence 15889, A	3283	70.4	38	10	US-09-899-651-8
Sequence 4296, Ap	3412	70.4	38	10	US-09-899-651-6
Sequence 8179, Ap	3667	70.4	38	10	US-10-097-340-146
Sequence 146, App	3694	70.4	38	10	US-09-764-864-329
Sequence 14943, A	400	69.4	37	5	US-09-960-352-14943
Sequence 3293, A	422	68.5	20	10	US-09-864-761-3293
Sequence 15889, A	557	68.5	37	10	US-09-864-761-15889
Sequence 4296, Ap	720	68.5	37	10	US-09-815-242-4296
Sequence 601, App	827	68.5	37	10	US-09-815-242-8179
Sequence 1844, Ap	1206	68.5	37	10	US-09-738-626-1844
Sequence 3053, Ap	2205	68.5	37	10	US-09-938-842A-3053
Sequence 458, App	2226	68.5	37	10	US-09-822-830A-458
Sequence 83, Appl	3151	68.5	37	10	US-10-114-893-83
Sequence 41, Appl	7149	68.5	37	10	US-10-174-363-41
Sequence 1, Appli	7149	68.5	37	10	US-10-196-935A-1
Sequence 8028, Ap	19167	68.5	37	10	US-09-764-891-8028
Sequence 30, Appl	51552	68.5	37	10	US-09-733-294A-30
Sequence 24491, A	191	66.7	36	10	US-09-864-761-24491
Sequence 10996, A	297	66.7	36	10	US-09-783-590-10996
Sequence 753, App	372	66.7	36	10	US-09-560-863-753
Sequence 3735, Ap	393	66.7	36	10	US-09-960-352-3735
Sequence 28787, A	434	66.7	36	10	US-09-918-995-28787
Sequence 9622, Ap	441	66.7	36	10	US-09-960-352-9622
Sequence 31053, Ap	455	66.7	36	10	US-09-918-995-31053
Sequence 32467, A	496	66.7	36	10	US-09-918-995-32467
Sequence 7787, Ap	529	66.7	36	10	US-09-864-761-7787
Sequence 1478, Ap	629	66.7	36	10	US-09-764-891-1478
Sequence 23, Appl	700	66.7	36	10	US-09-832-129-23
Sequence 1, Appli	1041	66.7	36	10	US-09-751-299-1
Sequence 45, Appl	1142	66.7	36	10	US-09-764-868-45
Sequence 13, Appl	1326	66.7	36	10	US-09-388-089B-13
Sequence 1, Appli	1347	66.7	36	10	US-09-388-089B-1

78	36	66.7	1374	10	US-09-070-927A-650	Sequence 650, App	c 151	35	64.8	730	9	US-10-017-754-143	Sequence 143, App
79	36	66.7	1395	10	US-09-388-089B-10	Sequence 10, Appl	c 152	35	64.8	738	9	US-10-112-267-38	Sequence 38, Appl
c 80	36	66.7	1902	9	US-09-738-626-1138	Sequence 1138, Ap	c 153	35	64.8	841	9	US-10-112-267-39	Sequence 39, Appl
81	36	66.7	1931	10	US-09-925-300-584	Sequence 584, App	c 154	35	64.8	871	9	US-10-228-794-18	Sequence 18, Appl
82	36	66.7	2370	9	US-09-712-363-98	Sequence 98, Appl	c 155	35	64.8	871	9	US-09-983-802-43	Sequence 43, Appl
83	36	66.7	2567	9	US-09-938-842A-1308	Sequence 1308, Ap	c 156	35	64.8	917	9	US-09-284-320-47	Sequence 47, Appl
84	36	66.7	2764	10	US-09-928-301-326	Sequence 326, App	c 157	35	64.8	990	9	US-09-815-242-7696	Sequence 7696, Ap
85	36	66.7	4828	10	US-10-080-503-16	Sequence 16, Appl	c 158	35	64.8	1164	10	US-10-066-500-38	Sequence 38, Appl
86	36	66.7	4947	10	US-09-118-276-21	Sequence 21, Appl	c 159	35	64.8	1215	9	US-10-002-796-38	Sequence 38, Appl
c 87	36	66.7	6854	9	US-09-922-683-7	Sequence 7, Appl	c 160	35	64.8	1215	9	US-10-066-273-38	Sequence 38, Appl
c 88	36	66.7	17397	10	US-10-091-504-1945	Sequence 1945, Ap	c 161	35	64.8	1215	9	US-10-066-494-38	Sequence 38, Appl
c 89	36	66.7	17458	9	US-10-055-001A-25	Sequence 25, Appl	c 162	35	64.8	1215	9	US-10-066-269-38	Sequence 38, Appl
90	36	66.7	17476	9	US-10-055-001A-24	Sequence 24, Appl	c 163	35	64.8	1215	9	US-10-066-193-38	Sequence 38, Appl
91	36	66.7	17681	9	US-10-055-001A-26	Sequence 23, Appl	c 164	35	64.8	1215	9	US-10-066-211-38	Sequence 38, Appl
92	36	66.7	17862	9	US-10-055-001A-23	Sequence 13, Appl	c 165	35	64.8	1224	9	US-09-981-876-22	Sequence 22, Appl
93	36	66.7	18691	9	US-10-055-001A-13	Sequence 1943, Ap	c 166	35	64.8	1224	9	US-09-148-545-22	Sequence 22, Appl
c 94	36	66.7	19334	9	US-10-091-504-1943	Sequence 1943, Ap	c 167	35	64.8	1238	7	US-08-781-986A-232	Sequence 232, App
c 95	36	66.7	19334	10	US-09-764-869-1943	Sequence 1944, Ap	c 168	35	64.8	1266	9	US-10-028-072-319	Sequence 319, App
c 96	36	66.7	19345	10	US-10-091-504-1944	Sequence 1944, Ap	c 169	35	64.8	1266	9	US-10-121-049-319	Sequence 319, App
c 97	36	66.7	19345	10	US-09-764-869-1944	Sequence 1733, Ap	c 170	35	64.8	1266	9	US-10-123-904-319	Sequence 319, App
c 98	36	66.7	21606	9	US-09-764-869-1733	Sequence 1733, Ap	c 171	35	64.8	1266	9	US-10-140-470-319	Sequence 319, App
c 99	36	66.7	21606	10	US-09-764-869-1733	Sequence 1733, Ap	c 172	35	64.8	1266	9	US-10-175-746-319	Sequence 319, App
c 100	36	66.7	21606	10	US-09-927-091-5	Sequence 5, Appl	c 173	35	64.8	1266	9	US-10-176-918-319	Sequence 319, App
c 101	36	66.7	30625	10	US-09-927-091-5	Sequence 4, Appl	c 174	35	64.8	1266	9	US-10-176-921-319	Sequence 319, App
c 102	36	66.7	513509	9	US-09-754-853A-4	Sequence 6883, Ap	c 175	35	64.8	1266	9	US-10-137-865-319	Sequence 319, App
c 103	35	64.8	188	10	US-09-867-701-6883	Sequence 31758, A	c 176	35	64.8	1266	9	US-10-140-474-319	Sequence 319, App
c 104	35	64.8	192	10	US-09-864-761-31758	Sequence 406, App	c 177	35	64.8	1266	9	US-10-142-431-319	Sequence 319, App
c 105	35	64.8	201	9	US-10-046-935-406	Sequence 406, App	c 178	35	64.8	1266	9	US-10-143-114-319	Sequence 319, App
c 106	35	64.8	201	9	US-09-878-178-406	Sequence 406, App	c 179	35	64.8	1266	9	US-10-140-002-319	Sequence 319, App
c 107	35	64.8	201	9	US-10-146-502-406	Sequence 21097, A	c 180	35	64.8	1266	9	US-10-142-419-319	Sequence 319, App
c 108	35	64.8	240	10	US-09-864-761-21097	Sequence 29769, A	c 181	35	64.8	1266	9	US-10-123-262-319	Sequence 319, App
c 109	35	64.8	247	10	US-09-864-761-29769	Sequence 989, App	c 182	35	64.8	1266	9	US-10-142-423-319	Sequence 319, App
c 110	35	64.8	258	10	US-09-923-876-989	Sequence 5089, Ap	c 183	35	64.8	1266	9	US-10-121-050-319	Sequence 319, App
c 111	35	64.8	279	10	US-09-294-093B-5089	Sequence 4593, Ap	c 184	35	64.8	1266	9	US-10-141-755-319	Sequence 319, App
c 112	35	64.8	289	10	US-09-923-876-4593	Sequence 4593, Ap	c 185	35	64.8	1266	9	US-10-143-032-319	Sequence 319, App
c 113	35	64.8	289	10	US-09-998-598-618	Sequence 618, App	c 186	35	64.8	1266	9	US-10-123-236-319	Sequence 319, App
c 114	35	64.8	306	10	US-09-864-761-31011	Sequence 31011, A	c 187	35	64.8	1266	9	US-10-123-261-319	Sequence 319, App
c 115	35	64.8	335	9	US-09-796-692-7185	Sequence 7185, Ap	c 188	35	64.8	1266	9	US-10-140-921-319	Sequence 319, App
c 116	35	64.8	335	9	US-10-040-862-7185	Sequence 7185, Ap	c 189	35	64.8	1266	9	US-10-140-928-319	Sequence 319, App
c 117	35	64.8	349	10	US-09-764-877-4029	Sequence 4029, Ap	c 190	35	64.8	1266	9	US-10-121-043-319	Sequence 319, App
c 118	35	64.8	367	10	US-09-880-107-3638	Sequence 3638, Ap	c 191	35	64.8	1266	9	US-10-123-903-319	Sequence 319, App
c 119	35	64.8	380	9	US-10-060-036-821	Sequence 821, App	c 192	35	64.8	1266	9	US-10-124-819-319	Sequence 319, App
c 120	35	64.8	389	10	US-09-983-965-3527	Sequence 3527, Ap	c 193	35	64.8	1266	9	US-10-140-925-319	Sequence 319, App
c 121	35	64.8	417	9	US-09-918-995-210	Sequence 210, App	c 194	35	64.8	1266	9	US-10-160-498-319	Sequence 319, App
c 122	35	64.8	417	10	US-09-867-701-1131	Sequence 1131, Ap	c 195	35	64.8	1266	9	US-10-121-041-319	Sequence 319, App
c 123	35	64.8	421	10	US-09-764-877-559	Sequence 559, App	c 196	35	64.8	1266	9	US-10-121-047-319	Sequence 319, App
c 124	35	64.8	425	9	US-09-729-658B-94	Sequence 94, Appl	c 197	35	64.8	1266	9	US-10-123-215-319	Sequence 319, App
c 125	35	64.8	449	10	US-09-864-761-4346	Sequence 4346, Ap	c 198	35	64.8	1266	9	US-10-123-902-319	Sequence 319, App
c 126	35	64.8	452	9	US-09-918-995-15760	Sequence 15760, A	c 199	35	64.8	1266	9	US-10-124-813-319	Sequence 319, App
c 127	35	64.8	452	10	US-09-864-761-14462	Sequence 14462, A	c 200	35	64.8	1266	9	US-10-124-824-319	Sequence 319, App
c 128	35	64.8	454	10	US-09-960-352-10804	Sequence 10804, A	c 201	35	64.8	1266	9	US-10-125-922-319	Sequence 319, App
c 129	35	64.8	466	9	US-09-918-995-21258	Sequence 21258, A	c 202	35	64.8	1266	9	US-10-127-829A-319	Sequence 319, App
c 130	35	64.8	479	9	US-09-918-995-23618	Sequence 23618, A	c 203	35	64.8	1266	9	US-10-127-835A-319	Sequence 319, App
c 131	35	64.8	482	10	US-09-864-761-15235	Sequence 15235, A	c 204	35	64.8	1266	9	US-10-127-901A-319	Sequence 319, App
c 132	35	64.8	489	9	US-09-918-995-4232	Sequence 4232, Ap	c 205	35	64.8	1266	9	US-10-128-693A-319	Sequence 319, App
c 133	35	64.8	492	9	US-09-918-995-1440	Sequence 1440, Ap	c 206	35	64.8	1266	9	US-10-131-813A-319	Sequence 319, App
c 134	35	64.8	493	9	US-09-918-995-29611	Sequence 29611, A	c 207	35	64.8	1266	9	US-10-131-818A-319	Sequence 319, App
c 135	35	64.8	501	9	US-09-918-995-20409	Sequence 20409, A	c 208	35	64.8	1266	9	US-10-131-823A-319	Sequence 319, App
c 136	35	64.8	509	10	US-09-864-761-13200	Sequence 13200, A	c 209	35	64.8	1266	9	US-10-131-830A-319	Sequence 319, App
c 137	35	64.8	529	10	US-09-864-761-13200	Sequence 14, Appl	c 210	35	64.8	1266	9	US-10-131-837A-319	Sequence 319, App
c 138	35	64.8	535	10	US-09-560-863-14	Sequence 14, Appl	c 211	35	64.8	1266	9	US-10-137-872A-319	Sequence 319, App
c 139	35	64.8	673	10	US-09-764-877-976	Sequence 976, App	c 212	35	64.8	1266	9	US-10-127-835A-319	Sequence 319, App
c 140	35	64.8	684	9	US-09-736-457-1441	Sequence 1441, Ap	c 213	35	64.8	1266	9	US-10-127-835A-319	Sequence 319, App
c 141	35	64.8	684	9	US-09-902-941-1441	Sequence 1441, Ap	c 214	35	64.8	1266	9	US-10-127-901A-319	Sequence 319, App
c 142	35	64.8	684	9	US-09-849-626-1441	Sequence 1441, Ap	c 215	35	64.8	1266	9	US-10-128-693A-319	Sequence 319, App
c 143	35	64.8	684	9	US-10-017-754-1441	Sequence 1441, Ap	c 216	35	64.8	1266	9	US-10-131-813A-319	Sequence 319, App
c 144	35	64.8	690	10	US-09-924-401-47	Sequence 47, Appl	c 217	35	64.8	1266	9	US-10-131-818A-319	Sequence 319, App
c 145	35	64.8	699	10	US-09-879-536-290	Sequence 250, App	c 218	35	64.8	1266	9	US-10-131-823A-319	Sequence 319, App
c 146	35	64.8	700	9	US-10-228-794-53	Sequence 53, Appl	c 219	35	64.8	1266	9	US-10-131-830A-319	Sequence 319, App
c 147	35	64.8	725	9	US-10-202-193-307	Sequence 307, App	c 220	35	64.8	1266	9	US-10-131-837A-319	Sequence 319, App
c 148	35	64.8	730	9	US-09-736-457-143	Sequence 143, App	c 221	35	64.8	1266	9	US-10-131-837A-319	Sequence 319, App
c 149	35	64.8	730	9	US-09-902-941-143	Sequence 143, App	c 222	35	64.8	1266	9	US-10-131-837A-319	Sequence 319, App
c 150	35	64.8	730	9	US-09-849-626-143	Sequence 143, App	c 223	35	64.8	1266	9	US-10-137-872A-319	Sequence 319, App

C 224	35	64.8	1266	9	US-10-140-860-319	Sequence 319, App	c 297	35	64.8	1266	9	US-10-131-822A-319	Sequence 319, App
C 225	35	64.8	1266	9	US-10-142-417-319	Sequence 319, App	C 298	35	64.8	1266	9	US-10-131-828A-319	Sequence 319, App
C 226	35	64.8	1266	9	US-10-147-500-319	Sequence 319, App	C 299	35	64.8	1266	9	US-10-131-835A-319	Sequence 319, App
C 227	35	64.8	1266	9	US-10-147-503-319	Sequence 319, App	C 300	35	64.8	1266	9	US-10-137-864A-319	Sequence 319, App
C 228	35	64.8	1266	9	US-10-147-515-319	Sequence 319, App	C 301	35	64.8	1266	9	US-10-137-869A-319	Sequence 319, App
C 229	35	64.8	1266	9	US-10-147-517-319	Sequence 319, App	C 302	35	64.8	1266	9	US-10-147-523-319	Sequence 319, App
C 230	35	64.8	1266	9	US-10-147-519-319	Sequence 319, App	C 303	35	64.8	1266	9	US-10-158-785-319	Sequence 319, App
C 231	35	64.8	1266	9	US-10-147-526-319	Sequence 319, App	C 304	35	64.8	1266	9	US-10-112-267-13	Sequence 13, Appl
C 232	35	64.8	1266	9	US-10-147-527-319	Sequence 319, App	C 305	35	64.8	1266	9	US-10-112-267-14	Sequence 14, Appl
C 233	35	64.8	1266	9	US-10-152-395-319	Sequence 319, App	C 306	35	64.8	1266	9	US-09-284-320-72	Sequence 72, Appl
C 234	35	64.8	1266	9	US-10-157-782-319	Sequence 319, App	C 307	35	64.8	1266	9	US-09-915-582-30	Sequence 30, Appl
C 235	35	64.8	1266	9	US-10-121-040-319	Sequence 319, App	C 308	35	64.8	1266	9	US-09-915-582-14	Sequence 14, Appl
C 236	35	64.8	1266	9	US-10-121-056-319	Sequence 319, App	C 309	35	64.8	1266	9	US-09-883-797-7	Sequence 7, Appl
C 237	35	64.8	1266	9	US-10-121-061-319	Sequence 319, App	C 310	35	64.8	1266	9	US-09-938-842A-2597	Sequence 2597, A
C 238	35	64.8	1266	9	US-10-123-235-319	Sequence 319, App	C 311	35	64.8	1266	9	US-09-864-761-18818	Sequence 18818, A
C 239	35	64.8	1266	9	US-10-124-818-319	Sequence 319, App	C 312	35	64.8	1266	9	US-09-752-385-7	Sequence 7, Appl
C 240	35	64.8	1266	9	US-10-125-926A-319	Sequence 319, App	C 313	35	64.8	1266	9	US-10-260-046-21	Sequence 21, Appl
C 241	35	64.8	1266	9	US-10-125-930A-319	Sequence 319, App	C 314	35	64.8	1266	9	US-10-152-661-588	Sequence 588, App
C 242	35	64.8	1266	9	US-10-127-831A-319	Sequence 319, App	C 315	35	64.8	1266	9	US-09-866-050A-588	Sequence 588, App
C 243	35	64.8	1266	9	US-10-127-837A-319	Sequence 319, App	C 316	35	64.8	1266	9	US-10-124-880-15	Sequence 15, Appl
C 244	35	64.8	1266	9	US-10-127-838B-319	Sequence 319, App	C 317	35	64.8	1266	9	US-09-764-898-76	Sequence 76, Appl
C 245	35	64.8	1266	9	US-10-127-842A-319	Sequence 319, App	C 318	35	64.8	1266	9	US-09-864-761-2074	Sequence 2074, Ap
C 246	35	64.8	1266	9	US-10-127-843A-319	Sequence 319, App	C 319	35	64.8	1266	9	US-10-091-438-44	Sequence 44, Appl
C 247	35	64.8	1266	9	US-10-127-845A-319	Sequence 319, App	C 320	35	64.8	1266	9	US-09-764-853-144	Sequence 144, App
C 248	35	64.8	1266	9	US-10-127-846A-319	Sequence 319, App	C 321	35	64.8	1266	9	US-09-893-519A-146	Sequence 146, App
C 249	35	64.8	1266	9	US-10-127-849A-319	Sequence 319, App	C 322	35	64.8	1266	9	US-09-905-657-1	Sequence 1, Appl
C 250	35	64.8	1266	9	US-10-127-849A-319	Sequence 319, App	C 323	35	64.8	1266	9	US-09-938-842A-911	Sequence 911, App
C 251	35	64.8	1266	9	US-10-127-850A-319	Sequence 319, App	C 324	35	64.8	1266	9	US-09-764-864-490	Sequence 490, App
C 252	35	64.8	1266	9	US-10-127-851A-319	Sequence 319, App	C 325	35	64.8	1266	9	US-09-764-864-31	Sequence 31, Appl
C 253	35	64.8	1266	9	US-10-128-684A-319	Sequence 319, App	C 326	35	64.8	1266	9	US-10-006-856A-53	Sequence 53, Appl
C 254	35	64.8	1266	9	US-10-128-686A-319	Sequence 319, App	C 327	35	64.8	1266	9	US-10-006-818A-53	Sequence 53, Appl
C 255	35	64.8	1266	9	US-10-128-690A-319	Sequence 319, App	C 328	35	64.8	1266	9	US-10-015-393A-53	Sequence 53, Appl
C 256	35	64.8	1266	9	US-10-128-691A-319	Sequence 319, App	C 329	35	64.8	1266	9	US-09-946-374-53	Sequence 53, Appl
C 257	35	64.8	1266	9	US-10-131-819A-319	Sequence 319, App	C 330	35	64.8	1266	9	US-10-012-121A-53	Sequence 53, Appl
C 258	35	64.8	1266	9	US-10-131-829A-319	Sequence 319, App	C 331	35	64.8	1266	9	US-10-015-869A-53	Sequence 53, Appl
C 259	35	64.8	1266	9	US-10-131-836A-319	Sequence 319, App	C 332	35	64.8	1266	9	US-10-006-116A-53	Sequence 53, Appl
C 260	35	64.8	1266	9	US-10-137-868-319	Sequence 319, App	C 333	35	64.8	1266	9	US-10-006-117A-53	Sequence 53, Appl
C 261	35	64.8	1266	9	US-10-146-729-319	Sequence 319, App	C 334	35	64.8	1266	9	US-10-013-913A-53	Sequence 53, Appl
C 262	35	64.8	1266	9	US-10-146-791-319	Sequence 319, App	C 335	35	64.8	1266	9	US-10-017-527A-53	Sequence 53, Appl
C 263	35	64.8	1266	9	US-10-147-484-319	Sequence 319, App	C 336	35	64.8	1266	9	US-10-007-194A-53	Sequence 53, Appl
C 264	35	64.8	1266	9	US-10-147-492-319	Sequence 319, App	C 337	35	64.8	1266	9	US-10-013-430A-53	Sequence 53, Appl
C 265	35	64.8	1266	9	US-10-147-508-319	Sequence 319, App	C 338	35	64.8	1266	9	US-09-764-891-5945	Sequence 5945, Ap
C 266	35	64.8	1266	9	US-10-147-512-319	Sequence 319, App	C 339	35	64.8	1266	9	US-09-764-891-5945	Sequence 5945, Ap
C 267	35	64.8	1266	9	US-10-158-782-319	Sequence 319, App	C 340	35	64.8	1266	9	US-10-091-504-1828	Sequence 1828, Ap
C 268	35	64.8	1266	9	US-10-175-735-319	Sequence 319, App	C 341	35	64.8	1266	9	US-09-729-658B-18	Sequence 18, Appl
C 269	35	64.8	1266	9	US-10-123-905-319	Sequence 319, App	C 342	35	64.8	1266	9	US-10-080-505-8	Sequence 8, Appl
C 270	35	64.8	1266	9	US-10-123-907-319	Sequence 319, App	C 343	35	64.8	1266	9	US-09-839-996-1	Sequence 1, Appl
C 271	35	64.8	1266	9	US-10-124-815-319	Sequence 319, App	C 344	35	64.8	1266	9	US-10-080-505-10	Sequence 10, Appl
C 272	35	64.8	1266	9	US-10-125-921A-319	Sequence 319, App	C 345	35	64.8	1266	9	US-10-080-505-14	Sequence 14, Appl
C 273	35	64.8	1266	9	US-10-125-928A-319	Sequence 319, App	C 346	35	64.8	1266	9	US-09-764-891-7855	Sequence 7855, Ap
C 274	35	64.8	1266	9	US-10-127-821A-319	Sequence 319, App	C 347	35	64.8	1266	9	US-10-080-505-12	Sequence 12, Appl
C 275	35	64.8	1266	9	US-10-127-822A-319	Sequence 319, App	C 348	35	64.8	1266	9	US-09-712-363-115	Sequence 115, App
C 276	35	64.8	1266	9	US-10-127-824A-319	Sequence 319, App	C 349	35	64.8	1266	9	US-09-566-880A-110	Sequence 110, Appl
C 277	35	64.8	1266	9	US-10-127-826A-319	Sequence 319, App	C 350	35	64.8	1266	9	US-09-764-891-8442	Sequence 8442, Ap
C 278	35	64.8	1266	9	US-10-127-827A-319	Sequence 319, App	C 351	35	64.8	1266	9	US-10-091-504-1821	Sequence 1821, Ap
C 279	35	64.8	1266	9	US-10-127-828A-319	Sequence 319, App	C 352	35	64.8	1266	9	US-09-764-891-7856	Sequence 7856, Ap
C 280	35	64.8	1266	9	US-10-127-830A-319	Sequence 319, App	C 353	35	64.8	1266	9	US-09-764-869-1821	Sequence 1821, Ap
C 281	35	64.8	1266	9	US-10-127-832A-319	Sequence 319, App	C 354	35	64.8	1266	9	US-10-044-090-253	Sequence 253, App
C 282	35	64.8	1266	9	US-10-127-833A-319	Sequence 319, App	C 355	35	64.8	1266	9	US-10-074-475-93	Sequence 93, Appl
C 283	35	64.8	1266	9	US-10-127-834A-319	Sequence 319, App	C 356	35	64.8	1266	9	US-10-091-504-1819	Sequence 1819, Ap
C 284	35	64.8	1266	9	US-10-127-836A-319	Sequence 319, App	C 357	35	64.8	1266	9	US-09-764-891-7854	Sequence 7854, Ap
C 285	35	64.8	1266	9	US-10-127-841A-319	Sequence 319, App	C 358	35	64.8	1266	9	US-09-764-869-1819	Sequence 1819, Ap
C 286	35	64.8	1266	9	US-10-127-844A-319	Sequence 319, App	C 359	35	64.8	1266	9	US-09-764-869-1819	Sequence 1819, Ap
C 287	35	64.8	1266	9	US-10-128-687A-319	Sequence 319, App	C 360	35	64.8	1266	9	US-09-764-891-7859	Sequence 7859, Ap
C 288	35	64.8	1266	9	US-10-128-688A-319	Sequence 319, App	C 361	35	64.8	1266	9	US-09-764-869-1826	Sequence 1826, Ap
C 289	35	64.8	1266	9	US-10-128-689A-319	Sequence 319, App	C 362	35	64.8	1266	9	US-09-764-869-1825	Sequence 1825, Ap
C 290	35	64.8	1266	9	US-10-128-694A-319	Sequence 319, App	C 363	35	64.8	1266	9	US-09-764-891-7858	Sequence 7858, Ap
C 291	35	64.8	1266	9	US-10-131-825A-319	Sequence 319, App	C 364	35	64.8	1266	9	US-09-764-869-1825	Sequence 1825, Ap
C 292	35	64.8	1266	9	US-10-230-417-319	Sequence 319, App	C 365	35	64.8	1266	9	US-09-764-869-1829	Sequence 1829, Ap
C 293	35	64.8	1266	9	US-10-121-051-319	Sequence 319, App	C 366	35	64.8	1266	9	US-10-091-504-1829	Sequence 7861, Ap
C 294	35	64.8	1266	9	US-10-131-815A-319	Sequence 319, App	C 367	35	64.8	1266	9	US-09-764-869-1829	Sequence 1829, Ap
C 295	35	64.8	1266	9	US-10-131-817A-319	Sequence 319, App	C 368	35	64.8	1266	9	US-09-764-869-1829	Sequence 1829, Ap
C 296	35	64.8	1266	9	US-10-131-821A-319	Sequence 319, App	C 369	35	64.8	1266	9	US-09-764-869-1829	Sequence 1829, Ap

370	35	64.8	9805	9	US-10-091-504-1824	Sequence 1824, Ap	443	34	63.0	486	10	US-09-974-300-1922	Sequence 1922, Ap
371	35	64.8	9805	9	US-09-764-891-7857	Sequence 7857, Ap	C 444	34	63.0	487	9	US-09-918-995-23	Sequence 23, Appl
372	35	64.8	9805	10	US-09-764-869-1824	Sequence 1824, Ap	C 445	34	63.0	489	9	US-09-918-995-30543	Sequence 30543, A
373	35	64.8	9831	10	US-10-091-504-1827	Sequence 1827, Ap	C 446	34	63.0	499	12	US-10-044-090-133	Sequence 133, App
374	35	64.8	9831	10	US-09-764-869-1827	Sequence 1827, Ap	C 447	34	63.0	501	9	US-09-918-995-32151	Sequence 32151, A
375	35	64.8	10594	10	US-09-070-927A-440	Sequence 440, Ap	C 448	34	63.0	505	9	US-09-918-995-32151	Sequence 32151, A
376	35	64.8	10642	9	US-10-224-414-3	Sequence 3, Appl	C 449	34	63.0	510	9	US-09-918-995-32151	Sequence 32151, A
377	35	64.8	11204	9	US-09-966-880A-35	Sequence 35, Appl	C 450	34	63.0	517	9	US-10-060-036-33	Sequence 33, Appl
378	35	64.8	13862	9	US-09-764-891-5477	Sequence 5477, Ap	C 451	34	63.0	518	10	US-09-998-598-692	Sequence 692, App
379	35	64.8	13862	9	US-09-764-891-10204	Sequence 10204, A	C 452	34	63.0	519	9	US-09-918-995-31117	Sequence 31117, App
380	35	64.8	16018	9	US-10-091-504-1384	Sequence 1384, Ap	C 453	34	63.0	519	9	US-09-918-995-32641	Sequence 32641, A
381	35	64.8	16018	10	US-09-764-869-1384	Sequence 1384, Ap	C 454	34	63.0	523	9	US-09-918-995-31638	Sequence 31638, A
382	35	64.8	16579	9	US-09-764-891-8443	Sequence 8443, Ap	C 455	34	63.0	554	9	US-09-918-995-32092	Sequence 32092, A
383	35	64.8	17335	10	US-10-092-154-1280	Sequence 1280, Ap	C 456	34	63.0	586	9	US-09-918-995-25894	Sequence 25894, A
384	35	64.8	17335	10	US-09-764-847-1280	Sequence 1280, Ap	C 457	34	63.0	587	12	US-10-001-870-87	Sequence 87, Appl
385	35	64.8	19882	9	US-10-092-154-1281	Sequence 1281, Ap	C 458	34	63.0	618	12	US-10-081-218-2	Sequence 2, Appl
386	35	64.8	19882	10	US-09-764-847-1281	Sequence 1281, Ap	C 459	34	63.0	701	10	US-09-529-063-24	Sequence 24, Appl
387	35	64.8	22400	9	US-10-091-504-1385	Sequence 1385, Ap	C 460	34	63.0	730	10	US-09-919-580-848	Sequence 848, App
388	35	64.8	22400	10	US-09-764-869-1385	Sequence 1385, Ap	C 461	34	63.0	735	9	US-09-738-626-2186	Sequence 2186, Ap
389	35	64.8	26798	10	US-09-880-107-3949	Sequence 3949, Ap	C 462	34	63.0	738	10	US-09-876-527-24	Sequence 24, Appl
390	35	64.8	157875	9	US-09-935-464-1	Sequence 1, Appl	C 463	34	63.0	745	9	US-10-001-887-67	Sequence 67, Appl
391	35	64.8	157875	9	US-10-125-835-1	Sequence 1, Appl	C 464	34	63.0	748	9	US-09-822-846-368	Sequence 368, App
392	35	64.8	249487	9	US-10-026-188-3	Sequence 3, Appl	C 465	34	63.0	755	12	US-10-021-509-1	Sequence 1, Appl
393	35	64.8	302250	10	US-09-962-833-154	Sequence 154, App	C 466	34	63.0	757	10	US-09-974-300-302	Sequence 302, App
394	35	64.8	536165	10	US-09-939-964-1	Sequence 1, Appl	C 467	34	63.0	757	10	US-09-974-300-471	Sequence 471, App
395	34.5	63.9	380	10	US-09-960-352-1519	Sequence 1519, Ap	C 468	34	63.0	769	12	US-10-044-593-85	Sequence 85, Appl
396	34.5	63.9	10906	10	US-09-956-004-63	Sequence 63, Appl	C 469	34	63.0	932	10	US-09-876-527-25	Sequence 25, Appl
397	34.5	63.9	23378	9	US-10-092-154-1514	Sequence 1514, Ap	C 470	34	63.0	978	9	US-09-769-787-353	Sequence 353, App
398	34.5	63.9	23378	10	US-09-764-847-1514	Sequence 1514, Ap	C 471	34	63.0	983	9	US-09-822-846-226	Sequence 226, App
399	34	63.0	136	10	US-09-783-590-5139	Sequence 5139, Ap	C 472	34	63.0	983	12	US-10-081-218-3	Sequence 3, Appl
400	34	63.0	192	9	US-10-095-672A-6	Sequence 6, Appl	C 473	34	63.0	993	12	US-10-081-218-4	Sequence 4, Appl
401	34	63.0	212	7	US-08-781-986A-1745	Sequence 1745, Ap	C 474	34	63.0	999	10	US-09-876-527-26	Sequence 26, Appl
402	34	63.0	213	10	US-09-974-300-7463	Sequence 7463, Ap	C 475	34	63.0	1011	10	US-09-876-527-27	Sequence 27, Appl
403	34	63.0	231	9	US-10-091-504-361	Sequence 361, App	C 476	34	63.0	1040	9	US-10-102-806-185	Sequence 185, App
404	34	63.0	231	10	US-09-764-869-361	Sequence 361, App	C 477	34	63.0	1047	9	US-10-050-704-12	Sequence 12, Appl
405	34	63.0	247	10	US-09-923-876-1820	Sequence 1820, App	C 478	34	63.0	1061	10	US-09-764-864-247	Sequence 247, App
406	34	63.0	250	10	US-09-960-352-14211	Sequence 14211, A	C 479	34	63.0	1064	9	US-10-001-887-68	Sequence 68, Appl
407	34	63.0	267	10	US-09-974-300-7923	Sequence 7923, Ap	C 480	34	63.0	1071	10	US-09-263-959-242	Sequence 242, App
408	34	63.0	274	10	US-09-294-093B-947	Sequence 947, App	C 481	34	63.0	1118	9	US-10-227-884-31	Sequence 31, Appl
409	34	63.0	307	9	US-10-079-623-268	Sequence 268, App	C 482	34	63.0	1118	9	US-10-230-163-31	Sequence 31, Appl
410	34	63.0	312	10	US-09-974-300-3854	Sequence 3854, Ap	C 483	34	63.0	1118	9	US-10-218-631-31	Sequence 31, Appl
411	34	63.0	313	10	US-09-216-393-32	Sequence 32, Appl	C 484	34	63.0	1118	9	US-10-230-338-31	Sequence 31, Appl
412	34	63.0	326	10	US-09-960-352-401	Sequence 401, App	C 485	34	63.0	1118	9	US-10-230-414-31	Sequence 31, Appl
413	34	63.0	328	9	US-09-803-719-183	Sequence 183, App	C 486	34	63.0	1118	9	US-10-216-159A-31	Sequence 31, Appl
414	34	63.0	341	9	US-09-918-995-19920	Sequence 19920, A	C 487	34	63.0	1118	9	US-10-218-849-31	Sequence 31, Appl
415	34	63.0	363	10	US-09-960-352-1068	Sequence 1068, Ap	C 488	34	63.0	1118	9	US-10-227-873-31	Sequence 31, Appl
416	34	63.0	380	9	US-10-095-672A-8	Sequence 8, Appl	C 489	34	63.0	1118	9	US-10-227-883-31	Sequence 31, Appl
417	34	63.0	381	10	US-09-867-701-8928	Sequence 8928, Ap	C 490	34	63.0	1118	9	US-10-219-076-31	Sequence 31, Appl
418	34	63.0	383	10	US-09-867-701-3893	Sequence 3893, Ap	C 491	34	63.0	1118	9	US-10-230-434-31	Sequence 31, Appl
419	34	63.0	396	9	US-10-101-464A-221	Sequence 221, App	C 492	34	63.0	1118	9	US-10-219-003-31	Sequence 31, Appl
420	34	63.0	396	10	US-09-864-761-4628	Sequence 4628, Ap	C 493	34	63.0	1118	9	US-10-219-075-31	Sequence 31, Appl
421	34	63.0	396	10	US-09-864-761-21370	Sequence 21370, A	C 494	34	63.0	1118	9	US-10-219-466-31	Sequence 31, Appl
422	34	63.0	398	10	US-09-960-352-9564	Sequence 9564, Ap	C 495	34	63.0	1118	9	US-10-219-466-31	Sequence 31, Appl
423	34	63.0	399	10	US-09-983-965-4819	Sequence 4819, Ap	C 496	34	63.0	1118	9	US-10-219-479-31	Sequence 31, Appl
424	34	63.0	410	9	US-10-095-672A-7	Sequence 7, Appl	C 497	34	63.0	1118	9	US-10-219-481-31	Sequence 31, Appl
425	34	63.0	412	9	US-09-854-133-689	Sequence 689, App	C 498	34	63.0	1118	9	US-10-230-260-31	Sequence 31, Appl
426	34	63.0	413	9	US-09-918-995-8672	Sequence 8672, Ap	C 499	34	63.0	1118	9	US-10-232-231-31	Sequence 31, Appl
427	34	63.0	423	9	US-10-060-036-299	Sequence 299, App	C 500	34	63.0	1118	9	US-10-232-233-31	Sequence 31, Appl
428	34	63.0	425	10	US-09-960-352-10001	Sequence 10001, A	C 501	34	63.0	1118	9	US-10-216-165-31	Sequence 31, Appl
429	34	63.0	434	10	US-09-960-352-4929	Sequence 4929, Ap	C 502	34	63.0	1118	9	US-10-218-956-31	Sequence 31, Appl
430	34	63.0	437	9	US-09-918-995-37031	Sequence 37031, A	C 503	34	63.0	1118	9	US-10-219-468-31	Sequence 31, Appl
431	34	63.0	440	9	US-10-066-543-3301	Sequence 3301, Ap	C 504	34	63.0	1118	9	US-10-219-478-31	Sequence 31, Appl
432	34	63.0	442	9	US-10-007-280A-132	Sequence 132, App	C 505	34	63.0	1118	9	US-10-219-536-31	Sequence 31, Appl
433	34	63.0	450	10	US-09-529-063-23	Sequence 23, Appl	C 506	34	63.0	1150	10	US-09-263-959-238	Sequence 238, App
434	34	63.0	460	10	US-09-864-761-6282	Sequence 6282, Ap	C 507	34	63.0	1150	10	US-09-263-959-241	Sequence 241, App
435	34	63.0	461	9	US-09-918-995-11980	Sequence 11980, A	C 508	34	63.0	1151	9	US-09-910-186A-25	Sequence 25, Appl
436	34	63.0	465	10	US-09-974-300-1124	Sequence 1124, Ap	C 509	34	63.0	1165	10	US-09-263-959-240	Sequence 240, App
437	34	63.0	471	10	US-09-998-598-2300	Sequence 2300, Ap	C 510	34	63.0	1185	10	US-09-815-242-6136	Sequence 6136, Ap
438	34	63.0	476	9	US-09-918-995-11257	Sequence 11257, A	C 511	34	63.0	1203	10	US-09-263-959-239	Sequence 239, App
439	34	63.0	477	10	US-09-864-761-4431	Sequence 4431, Ap	C 512	34	63.0	1241	12	US-10-044-090-142	Sequence 142, App
440	34	63.0	480	9	US-09-918-995-7413	Sequence 7413, Ap	C 513	34	63.0	1242	10	US-09-925-301-159	Sequence 159, App
441	34	63.0	485	9	US-09-918-995-22805	Sequence 22805, A	C 514	34	63.0	1248	10	US-09-815-243-8495	Sequence 8495, Ap
442	34	63.0	485	9	US-09-918-995-28562	Sequence 28562, A	C 515	34	63.0	1248	10	US-09-815-242-8811	Sequence 8811, Ap

516	34	63.0	1249	9	US-09-774-381-48	Sequence 48, Appl	c 589	34	63.0	2475	9	US-10-011-588-32	Sequence 32, Appl
517	34	63.0	1320	9	US-09-854-133-22	Sequence 22, Appl	590	34	63.0	2718	9	US-09-989-919-11	Sequence 11, Appl
518	34	63.0	1320	10	US-09-738-973-22	Sequence 22, Appl	591	34	63.0	2736	10	US-09-876-527-23	Sequence 23, Appl
519	34	63.0	1328	9	US-09-854-133-24	Sequence 24, Appl	592	34	63.0	2904	9	US-10-012-896-703	Sequence 703, Appl
520	34	63.0	1328	10	US-09-738-973-24	Sequence 24, Appl	593	34	63.0	2904	9	US-09-895-793-703	Sequence 703, Appl
521	34	63.0	1331	9	US-09-854-133-27	Sequence 27, Appl	594	34	63.0	2904	9	US-09-895-814-703	Sequence 703, Appl
522	34	63.0	1331	10	US-09-738-973-27	Sequence 27, Appl	595	34	63.0	2904	10	US-09-759-143-703	Sequence 703, Appl
523	34	63.0	1333	9	US-09-854-133-28	Sequence 28, Appl	596	34	63.0	2904	10	US-09-780-669-703	Sequence 703, Appl
524	34	63.0	1333	10	US-09-738-973-28	Sequence 28, Appl	597	34	63.0	2904	10	US-09-822-827-703	Sequence 703, Appl
525	34	63.0	1335	9	US-09-854-133-31	Sequence 31, Appl	598	34	63.0	2968	10	US-09-822-827-703	Sequence 703, Appl
526	34	63.0	1355	9	US-09-738-973-31	Sequence 31, Appl	599	34	63.0	3153	9	US-09-815-915-1	Sequence 1, Appl
527	34	63.0	1361	9	US-09-925-299-72	Sequence 72, Appl	600	34	63.0	3153	9	US-09-759-1308-343	Sequence 343, Appl
528	34	63.0	1361	10	US-09-925-299-72	Sequence 72, Appl	601	34	63.0	3181	9	US-10-189-123-73	Sequence 73, Appl
529	34	63.0	1375	9	US-09-263-959-289	Sequence 289, Appl	602	34	63.0	3287	10	US-10-098-841-282	Sequence 282, Appl
530	34	63.0	1422	9	US-09-938-842A-2680	Sequence 2680, Ap	603	34	63.0	3320	10	US-09-876-527-15	Sequence 15, Appl
531	34	63.0	1461	10	US-09-815-243-4454	Sequence 4454, Ap	604	34	63.0	3320	10	US-09-838-785-1	Sequence 1, Appl
532	34	63.0	1521	9	US-09-738-626-2225	Sequence 2225, Ap	605	34	63.0	3348	10	US-09-312-762A-2	Sequence 2, Appl
533	34	63.0	1534	10	US-09-765-231A-20	Sequence 20, Appl	606	34	63.0	3410	9	US-09-232-880-110	Sequence 110, Appl
534	34	63.0	1542	10	US-09-974-300-2025	Sequence 2025, Ap	607	34	63.0	3410	9	US-10-012-896-110	Sequence 110, Appl
535	34	63.0	1566	10	US-09-815-242-7758	Sequence 7758, Ap	608	34	63.0	3410	9	US-09-895-793-110	Sequence 110, Appl
536	34	63.0	1743	10	US-09-917-800A-1349	Sequence 1349, Ap	609	34	63.0	3410	9	US-09-895-814-110	Sequence 110, Appl
537	34	63.0	1746	9	US-09-989-442-16	Sequence 16, Appl	610	34	63.0	3410	10	US-10-010-940-110	Sequence 110, Appl
538	34	63.0	1750	9	US-09-854-133-25	Sequence 25, Appl	611	34	63.0	3410	10	US-09-745-288-100	Sequence 100, Appl
539	34	63.0	1758	9	US-09-989-442-16	Sequence 16, Appl	612	34	63.0	3410	10	US-09-759-143-110	Sequence 110, Appl
540	34	63.0	1758	10	US-09-738-973-25	Sequence 25, Appl	613	34	63.0	3410	10	US-09-780-669-110	Sequence 110, Appl
541	34	63.0	1774	9	US-10-168-066-20	Sequence 20, Appl	614	34	63.0	3410	10	US-09-030-606-110	Sequence 110, Appl
542	34	63.0	1809	9	US-10-124-429-3	Sequence 3, Appl	615	34	63.0	3410	10	US-09-822-827-110	Sequence 110, Appl
543	34	63.0	1809	9	US-10-112-488-12	Sequence 12, Appl	616	34	63.0	3453	9	US-09-115-453-110	Sequence 110, Appl
544	34	63.0	1965	10	US-09-880-107-3843	Sequence 3843, Ap	617	34	63.0	4026	10	US-10-101-464A-861	Sequence 861, Appl
545	34	63.0	2003	10	US-09-887-576-319	Sequence 319, Appl	618	34	63.0	4034	9	US-09-736-960-3	Sequence 3, Appl
546	34	63.0	2025	9	US-10-173-123-8	Sequence 8, Appl	619	34	63.0	4034	9	US-10-012-896-704	Sequence 704, Appl
547	34	63.0	2042	9	US-09-984-245-21	Sequence 21, Appl	620	34	63.0	4034	9	US-09-895-793-704	Sequence 704, Appl
548	34	63.0	2042	9	US-09-966-262-21	Sequence 21, Appl	621	34	63.0	4034	10	US-09-895-814-704	Sequence 704, Appl
549	34	63.0	2042	9	US-09-983-966-21	Sequence 21, Appl	622	34	63.0	4034	10	US-09-759-143-704	Sequence 704, Appl
550	34	63.0	2042	9	US-10-143-090-21	Sequence 21, Appl	623	34	63.0	4034	10	US-09-780-669-704	Sequence 704, Appl
551	34	63.0	2046	9	US-10-173-123-6	Sequence 6, Appl	624	34	63.0	4053	9	US-09-822-827-704	Sequence 704, Appl
552	34	63.0	2065	9	US-10-245-103-49	Sequence 49, Appl	625	34	63.0	4053	9	US-09-905-291A-293	Sequence 293, Appl
553	34	63.0	2065	9	US-10-245-107-49	Sequence 49, Appl	626	34	63.0	4053	9	US-09-902-853-293	Sequence 293, Appl
554	34	63.0	2065	9	US-10-245-143-49	Sequence 49, Appl	627	34	63.0	4053	9	US-09-907-841-293	Sequence 293, Appl
555	34	63.0	2065	9	US-10-245-771-49	Sequence 49, Appl	628	34	63.0	4053	9	US-09-907-841-293	Sequence 293, Appl
556	34	63.0	2065	9	US-10-245-551-49	Sequence 49, Appl	629	34	63.0	4053	9	US-09-904-011-293	Sequence 293, Appl
557	34	63.0	2065	9	US-10-245-883-49	Sequence 49, Appl	630	34	63.0	4053	9	US-10-028-072-351	Sequence 351, Appl
558	34	63.0	2065	9	US-10-237-535-49	Sequence 49, Appl	631	34	63.0	4053	9	US-09-906-742-293	Sequence 293, Appl
559	34	63.0	2065	9	US-10-238-183-49	Sequence 49, Appl	632	34	63.0	4053	9	US-10-121-049-351	Sequence 351, Appl
560	34	63.0	2065	9	US-10-238-283-49	Sequence 49, Appl	633	34	63.0	4053	9	US-10-123-904-351	Sequence 351, Appl
561	34	63.0	2065	9	US-10-238-370-49	Sequence 49, Appl	634	34	63.0	4053	9	US-10-140-470-351	Sequence 351, Appl
562	34	63.0	2065	9	US-10-245-055-49	Sequence 49, Appl	635	34	63.0	4053	9	US-09-906-838-293	Sequence 293, Appl
563	34	63.0	2065	9	US-10-245-147-49	Sequence 49, Appl	636	34	63.0	4053	9	US-09-907-613-293	Sequence 293, Appl
564	34	63.0	2065	9	US-10-245-730-49	Sequence 49, Appl	637	34	63.0	4053	9	US-09-907-942-293	Sequence 293, Appl
565	34	63.0	2065	9	US-10-245-739-49	Sequence 49, Appl	638	34	63.0	4053	9	US-10-175-746-351	Sequence 351, Appl
566	34	63.0	2065	9	US-10-246-210-49	Sequence 49, Appl	639	34	63.0	4053	9	US-10-176-918-351	Sequence 351, Appl
567	34	63.0	2065	9	US-10-239-196-49	Sequence 49, Appl	640	34	63.0	4053	9	US-10-176-921-351	Sequence 351, Appl
568	34	63.0	2065	9	US-10-243-024-49	Sequence 49, Appl	641	34	63.0	4053	9	US-10-137-865-351	Sequence 351, Appl
569	34	63.0	2065	9	US-10-243-409-49	Sequence 49, Appl	642	34	63.0	4053	9	US-10-140-474-351	Sequence 351, Appl
570	34	63.0	2065	9	US-10-245-033-49	Sequence 49, Appl	643	34	63.0	4053	9	US-09-904-820-293	Sequence 293, Appl
571	34	63.0	2065	9	US-10-245-621-49	Sequence 49, Appl	644	34	63.0	4053	9	US-09-904-859-293	Sequence 293, Appl
572	34	63.0	2065	9	US-10-245-880-49	Sequence 49, Appl	645	34	63.0	4053	9	US-09-909-204-293	Sequence 293, Appl
573	34	63.0	2065	9	US-10-243-095-49	Sequence 49, Appl	646	34	63.0	4053	9	US-10-142-431-351	Sequence 351, Appl
574	34	63.0	2065	9	US-10-245-185-49	Sequence 49, Appl	647	34	63.0	4053	9	US-10-143-114-351	Sequence 351, Appl
575	34	63.0	2065	9	US-10-245-427-49	Sequence 49, Appl	648	34	63.0	4053	9	US-09-904-786-293	Sequence 293, Appl
576	34	63.0	2065	9	US-10-245-473-49	Sequence 49, Appl	649	34	63.0	4053	9	US-09-906-846-293	Sequence 293, Appl
577	34	63.0	2065	9	US-10-245-770-49	Sequence 49, Appl	650	34	63.0	4053	9	US-09-906-700-293	Sequence 293, Appl
578	34	63.0	2065	9	US-10-245-877-49	Sequence 49, Appl	651	34	63.0	4053	9	US-10-140-002-351	Sequence 351, Appl
579	34	63.0	2065	9	US-10-246-976-49	Sequence 49, Appl	652	34	63.0	4053	9	US-09-902-903-293	Sequence 293, Appl
580	34	63.0	2065	9	US-10-243-320-49	Sequence 49, Appl	653	34	63.0	4053	9	US-09-903-749A-293	Sequence 293, Appl
581	34	63.0	2158	8	US-10-023-515-1	Sequence 1, Appl	654	34	63.0	4053	9	US-09-903-786-293	Sequence 293, Appl
582	34	63.0	2172	10	US-08-908-884-13	Sequence 13, Appl	655	34	63.0	4053	9	US-10-142-419-351	Sequence 351, Appl
583	34	63.0	2217	10	US-09-908-323-13	Sequence 13, Appl	656	34	63.0	4053	9	US-09-902-736-293	Sequence 293, Appl
584	34	63.0	2218	9	US-10-173-123-12	Sequence 12, Appl	657	34	63.0	4053	9	US-09-904-119-293	Sequence 293, Appl
585	34	63.0	2238	9	US-10-173-123-10	Sequence 10, Appl	658	34	63.0	4053	9	US-09-904-956-293	Sequence 293, Appl
586	34	63.0	2287	9	US-10-152-661-543	Sequence 543, Appl	659	34	63.0	4053	9	US-09-907-794-293	Sequence 293, Appl
587	34	63.0	2287	9	US-09-866-050A-543	Sequence 543, Appl	660	34	63.0	4053	9	US-10-123-262-351	Sequence 351, Appl
588	34	63.0	2352	10	US-09-815-915-3	Sequence 3, Appl	661	34	63.0	4053	9	US-10-142-423-351	Sequence 351, Appl

662	34	63.0	4053	9	US-09-903-520-293	Sequence 293, App	735	34	63.0	4053	9	US-10-125-930A-351	Sequence 351, App
663	34	63.0	4053	9	US-09-903-943-293	Sequence 293, App	736	34	63.0	4053	9	US-10-127-831A-351	Sequence 351, App
664	34	63.0	4053	9	US-09-904-462-293	Sequence 293, App	737	34	63.0	4053	9	US-10-127-837A-351	Sequence 351, App
665	34	63.0	4053	9	US-09-905-056-293	Sequence 293, App	738	34	63.0	4053	9	US-10-127-838B-351	Sequence 351, App
666	34	63.0	4053	9	US-09-907-925-293	Sequence 293, App	739	34	63.0	4053	9	US-10-127-842A-351	Sequence 351, App
667	34	63.0	4053	9	US-10-121-050-351	Sequence 351, App	740	34	63.0	4053	9	US-10-127-843A-351	Sequence 351, App
668	34	63.0	4053	9	US-10-141-755-351	Sequence 351, App	741	34	63.0	4053	9	US-10-127-845A-351	Sequence 351, App
669	34	63.0	4053	9	US-09-904-553-293	Sequence 293, App	742	34	63.0	4053	9	US-10-127-846A-351	Sequence 351, App
670	34	63.0	4053	9	US-09-905-381-293	Sequence 293, App	743	34	63.0	4053	9	US-10-127-848A-351	Sequence 351, App
671	34	63.0	4053	9	US-09-909-064-293	Sequence 293, App	744	34	63.0	4053	9	US-10-127-849A-351	Sequence 351, App
672	34	63.0	4053	9	US-10-143-032-351	Sequence 351, App	745	34	63.0	4053	9	US-10-127-850A-351	Sequence 351, App
673	34	63.0	4053	9	US-10-123-108-351	Sequence 351, App	746	34	63.0	4053	9	US-10-127-851A-351	Sequence 351, App
674	34	63.0	4053	9	US-10-123-236-351	Sequence 351, App	747	34	63.0	4053	9	US-10-128-684A-351	Sequence 351, App
675	34	63.0	4053	9	US-10-123-261-351	Sequence 351, App	748	34	63.0	4053	9	US-10-128-686A-351	Sequence 351, App
676	34	63.0	4053	9	US-10-140-921-351	Sequence 351, App	749	34	63.0	4053	9	US-10-128-690A-351	Sequence 351, App
677	34	63.0	4053	9	US-10-140-928-351	Sequence 351, App	750	34	63.0	4053	9	US-10-128-691A-351	Sequence 351, App
678	34	63.0	4053	9	US-09-905-088-293	Sequence 293, App	751	34	63.0	4053	9	US-10-131-819A-351	Sequence 351, App
679	34	63.0	4053	9	US-09-907-575-293	Sequence 293, App	752	34	63.0	4053	9	US-10-131-829A-351	Sequence 351, App
680	34	63.0	4053	9	US-10-121-045-351	Sequence 351, App	753	34	63.0	4053	9	US-10-131-836A-351	Sequence 351, App
681	34	63.0	4053	9	US-10-123-292-351	Sequence 351, App	754	34	63.0	4053	9	US-10-137-868-351	Sequence 351, App
682	34	63.0	4053	9	US-10-123-903-351	Sequence 351, App	755	34	63.0	4053	9	US-10-146-729-351	Sequence 351, App
683	34	63.0	4053	9	US-10-124-819-351	Sequence 351, App	756	34	63.0	4053	9	US-10-146-791-351	Sequence 351, App
684	34	63.0	4053	9	US-10-124-822-351	Sequence 351, App	757	34	63.0	4053	9	US-10-147-484-351	Sequence 351, App
685	34	63.0	4053	9	US-10-140-925-351	Sequence 351, App	758	34	63.0	4053	9	US-10-147-492-351	Sequence 351, App
686	34	63.0	4053	9	US-10-160-498-351	Sequence 351, App	759	34	63.0	4053	9	US-10-147-508-351	Sequence 351, App
687	34	63.0	4053	9	US-09-902-759-293	Sequence 293, App	760	34	63.0	4053	9	US-10-147-512-351	Sequence 351, App
688	34	63.0	4053	9	US-09-905-075-293	Sequence 293, App	761	34	63.0	4053	9	US-10-158-782-351	Sequence 351, App
689	34	63.0	4053	9	US-10-121-041-351	Sequence 351, App	762	34	63.0	4053	9	US-10-175-735-351	Sequence 351, App
690	34	63.0	4053	9	US-10-121-043-351	Sequence 351, App	763	34	63.0	4053	9	US-10-123-905-351	Sequence 351, App
691	34	63.0	4053	9	US-10-121-047-351	Sequence 351, App	764	34	63.0	4053	9	US-10-123-907-351	Sequence 351, App
692	34	63.0	4053	9	US-10-123-215-351	Sequence 351, App	765	34	63.0	4053	9	US-10-124-815-351	Sequence 351, App
693	34	63.0	4053	9	US-10-123-902-351	Sequence 351, App	766	34	63.0	4053	9	US-10-125-921A-351	Sequence 351, App
694	34	63.0	4053	9	US-10-123-908-351	Sequence 351, App	767	34	63.0	4053	9	US-10-125-928A-351	Sequence 351, App
695	34	63.0	4053	9	US-10-123-909-351	Sequence 351, App	768	34	63.0	4053	9	US-10-127-821A-351	Sequence 351, App
696	34	63.0	4053	9	US-10-123-910-351	Sequence 351, App	769	34	63.0	4053	9	US-10-127-822A-351	Sequence 351, App
697	34	63.0	4053	9	US-10-124-813-351	Sequence 351, App	770	34	63.0	4053	9	US-10-127-824A-351	Sequence 351, App
698	34	63.0	4053	9	US-10-124-817-351	Sequence 351, App	771	34	63.0	4053	9	US-10-127-826A-351	Sequence 351, App
699	34	63.0	4053	9	US-10-124-824-351	Sequence 351, App	772	34	63.0	4053	9	US-10-127-827A-351	Sequence 351, App
700	34	63.0	4053	9	US-10-125-922-351	Sequence 351, App	773	34	63.0	4053	9	US-10-127-828A-351	Sequence 351, App
701	34	63.0	4053	9	US-10-125-924-351	Sequence 351, App	774	34	63.0	4053	9	US-10-127-830A-351	Sequence 351, App
702	34	63.0	4053	9	US-10-127-825A-351	Sequence 351, App	775	34	63.0	4053	9	US-10-127-832A-351	Sequence 351, App
703	34	63.0	4053	9	US-10-127-829A-351	Sequence 351, App	776	34	63.0	4053	9	US-10-127-833A-351	Sequence 351, App
704	34	63.0	4053	9	US-10-127-835A-351	Sequence 351, App	777	34	63.0	4053	9	US-10-127-834A-351	Sequence 351, App
705	34	63.0	4053	9	US-10-127-839A-351	Sequence 351, App	778	34	63.0	4053	9	US-10-127-836A-351	Sequence 351, App
706	34	63.0	4053	9	US-10-127-901A-351	Sequence 351, App	779	34	63.0	4053	9	US-10-127-841A-351	Sequence 351, App
707	34	63.0	4053	9	US-10-128-693A-351	Sequence 351, App	780	34	63.0	4053	9	US-10-127-844A-351	Sequence 351, App
708	34	63.0	4053	9	US-10-131-813A-351	Sequence 351, App	781	34	63.0	4053	9	US-10-128-687A-351	Sequence 351, App
709	34	63.0	4053	9	US-10-131-818A-351	Sequence 351, App	782	34	63.0	4053	9	US-10-128-688A-351	Sequence 351, App
710	34	63.0	4053	9	US-10-131-823A-351	Sequence 351, App	783	34	63.0	4053	9	US-10-128-689A-351	Sequence 351, App
711	34	63.0	4053	9	US-10-131-824A-351	Sequence 351, App	784	34	63.0	4053	9	US-10-128-694A-351	Sequence 351, App
712	34	63.0	4053	9	US-10-131-830A-351	Sequence 351, App	785	34	63.0	4053	9	US-10-131-825A-351	Sequence 351, App
713	34	63.0	4053	9	US-10-131-837A-351	Sequence 351, App	786	34	63.0	4053	9	US-10-130-417-351	Sequence 351, App
714	34	63.0	4053	9	US-10-137-872A-351	Sequence 351, App	787	34	63.0	4053	9	US-09-902-615-293	Sequence 293, App
715	34	63.0	4053	9	US-10-140-860-351	Sequence 351, App	788	34	63.0	4053	9	US-10-121-051-351	Sequence 351, App
716	34	63.0	4053	9	US-10-142-417-351	Sequence 351, App	789	34	63.0	4053	9	US-10-131-815A-351	Sequence 351, App
717	34	63.0	4053	9	US-10-147-500-351	Sequence 351, App	790	34	63.0	4053	9	US-10-131-817A-351	Sequence 351, App
718	34	63.0	4053	9	US-10-147-502-351	Sequence 351, App	791	34	63.0	4053	9	US-10-131-821A-351	Sequence 351, App
719	34	63.0	4053	9	US-10-147-515-351	Sequence 351, App	792	34	63.0	4053	9	US-10-131-822A-351	Sequence 351, App
720	34	63.0	4053	9	US-10-147-517-351	Sequence 351, App	793	34	63.0	4053	9	US-10-131-828A-351	Sequence 351, App
721	34	63.0	4053	9	US-10-147-519-351	Sequence 351, App	794	34	63.0	4053	9	US-10-131-835A-351	Sequence 351, App
722	34	63.0	4053	9	US-10-147-526-351	Sequence 351, App	795	34	63.0	4053	9	US-10-137-864A-351	Sequence 351, App
723	34	63.0	4053	9	US-10-147-527-351	Sequence 351, App	796	34	63.0	4053	9	US-10-137-869A-351	Sequence 351, App
724	34	63.0	4053	9	US-10-152-395-351	Sequence 351, App	797	34	63.0	4053	9	US-10-147-523-351	Sequence 351, App
725	34	63.0	4053	9	US-10-157-782-351	Sequence 351, App	798	34	63.0	4053	9	US-10-158-785-351	Sequence 351, App
726	34	63.0	4053	9	US-09-902-634-293	Sequence 293, App	799	34	63.0	4053	10	US-09-909-320-293	Sequence 293, App
727	34	63.0	4053	9	US-09-902-713-293	Sequence 293, App	800	34	63.0	4053	10	US-09-909-088B-293	Sequence 293, App
728	34	63.0	4053	9	US-09-907-979-293	Sequence 293, App	801	34	63.0	4555	10	US-09-880-107-3738	Sequence 3738, App
729	34	63.0	4053	9	US-10-121-040-351	Sequence 351, App	c 802	34	63.0	4894	9	US-10-012-896-702	Sequence 702, App
730	34	63.0	4053	9	US-10-121-056-351	Sequence 351, App	c 803	34	63.0	4894	9	US-09-895-793-702	Sequence 702, App
731	34	63.0	4053	9	US-10-121-061-351	Sequence 351, App	c 804	34	63.0	4894	9	US-09-895-814-702	Sequence 702, App
732	34	63.0	4053	9	US-10-123-235-351	Sequence 351, App	c 805	34	63.0	4894	10	US-09-759-143-702	Sequence 702, App
733	34	63.0	4053	9	US-10-124-818-351	Sequence 351, App	c 806	34	63.0	4894	10	US-09-780-669-702	Sequence 702, App
734	34	63.0	4053	9	US-10-125-926A-351	Sequence 351, App	c 807	34	63.0	4894	10	US-09-822-821-702	Sequence 702, App



808	34	63.0	5248	9	US-09-822-846-76	Sequence 76, Appl	881	33	61.1	369	10	US-09-770-791-379	Sequence 379, App
809	34	63.0	5434	9	US-09-954-531-172	Sequence 172, App	c 882	33	61.1	379	10	US-09-960-352-4996	Sequence 4996, App
810	34	63.0	5434	9	US-09-954-531-381	Sequence 381, App	c 883	33	61.1	380	9	US-09-918-995-6208	Sequence 6208, App
811	34	63.0	5434	9	US-10-144-577-1	Sequence 1, Appli	c 884	33	61.1	383	9	US-09-918-995-36973	Sequence 36973, A
812	34	63.0	5434	10	US-09-962-436-285	Sequence 285, App	c 885	33	61.1	394	10	US-09-864-761-10318	Sequence 10318, A
813	34	63.0	6383	9	US-10-092-154-1415	Sequence 1415, Ap	c 886	33	61.1	396	9	US-09-970-966-51	Sequence 51, Appl
814	34	63.0	6383	10	US-09-764-847-1415	Sequence 1415, Ap	c 887	33	61.1	396	10	US-09-825-294-51	Sequence 51, Appl
815	34	63.0	6976	9	US-10-012-896-705	Sequence 705, App	c 888	33	61.1	396	10	US-09-867-550-841	Sequence 841, App
816	34	63.0	6976	9	US-09-895-793-705	Sequence 705, App	c 889	33	61.1	397	10	US-09-864-761-1482	Sequence 1482, App
817	34	63.0	6976	9	US-09-895-814-705	Sequence 705, App	c 890	33	61.1	397	10	US-09-960-352-1421	Sequence 1421, App
818	34	63.0	6976	10	US-09-759-143-705	Sequence 705, App	c 891	33	61.1	401	10	US-09-960-352-5973	Sequence 5973, App
819	34	63.0	6976	10	US-09-780-669-705	Sequence 705, App	c 892	33	61.1	402	10	US-09-960-352-9253	Sequence 9253, App
820	34	63.0	6976	10	US-09-822-827-705	Sequence 705, App	c 893	33	61.1	402	10	US-09-960-352-14467	Sequence 14467, A
821	34	63.0	8036	10	US-09-824-864-1621	Sequence 1621, Ap	c 894	33	61.1	404	10	US-09-983-965-2962	Sequence 2962, App
822	34	63.0	8577	9	US-10-108-605-248	Sequence 248, App	c 895	33	61.1	406	9	US-09-918-995-16596	Sequence 16596, A
823	34	63.0	8850	12	US-10-093-684-1	Sequence 1, Appli	c 896	33	61.1	408	9	US-09-918-995-34099	Sequence 34099, A
824	34	63.0	9662	9	US-09-764-891-9774	Sequence 9774, Ap	c 897	33	61.1	410	10	US-09-960-352-3527	Sequence 3527, App
825	34	63.0	12368	10	US-09-956-004-20	Sequence 20, Appl	c 898	33	61.1	411	10	US-09-983-965-1011	Sequence 1011, App
826	34	63.0	13021	9	US-10-103-313-612	Sequence 612, App	c 899	33	61.1	415	10	US-09-783-590-7397	Sequence 7397, App
827	34	63.0	13021	9	US-10-115-928-55	Sequence 55, Appl	c 900	33	61.1	415	10	US-09-960-352-7443	Sequence 7443, App
828	34	63.0	14707	10	US-09-312-762A-3	Sequence 3, Appli	c 901	33	61.1	419	9	US-09-918-995-35639	Sequence 35639, A
829	34	63.0	14792	9	US-09-764-891-9780	Sequence 9780, Ap	c 902	33	61.1	423	9	US-09-764-891-971	Sequence 971, App
830	34	63.0	16592	7	US-08-781-986A-53	Sequence 53, Appl	c 903	33	61.1	423	10	US-09-960-352-5959	Sequence 5959, App
831	34	63.0	16851	9	US-09-764-891-9781	Sequence 9781, Ap	c 904	33	61.1	423	10	US-09-960-352-11861	Sequence 11861, A
832	34	63.0	16853	9	US-09-764-891-9782	Sequence 9782, Ap	c 905	33	61.1	425	10	US-09-864-761-10454	Sequence 10454, A
833	34	63.0	21732	9	US-09-764-891-9773	Sequence 9773, Ap	c 906	33	61.1	427	10	US-09-983-965-953	Sequence 953, App
834	34	63.0	21732	9	US-10-092-154-1414	Sequence 1414, Ap	c 907	33	61.1	428	9	US-09-918-995-2153	Sequence 2153, App
835	34	63.0	21732	10	US-09-764-847-1414	Sequence 1414, Ap	c 908	33	61.1	429	9	US-09-954-531-946	Sequence 946, App
836	34	63.0	23106	9	US-09-863-049A-1	Sequence 1, Appli	c 909	33	61.1	429	9	US-09-954-531-1341	Sequence 1341, App
837	34	63.0	29729	10	US-09-070-927A-238	Sequence 238, App	c 910	33	61.1	429	10	US-09-954-456-937	Sequence 937, App
838	34	63.0	53522	9	US-09-904-968A-1	Sequence 1, Appli	c 911	33	61.1	429	10	US-09-954-456-1579	Sequence 1579, App
839	34	63.0	62804	12	US-10-096-960-3	Sequence 3, Appli	c 912	33	61.1	429	10	US-09-960-352-1407	Sequence 1407, App
840	34	63.0	66686	10	US-09-736-960-86	Sequence 86, Appl	c 913	33	61.1	429	10	US-09-880-107-3594	Sequence 3594, App
841	34	63.0	80959	9	US-09-858-546-3	Sequence 3, Appli	c 914	33	61.1	429	10	US-09-974-300-867	Sequence 867, App
842	34	63.0	81940	9	US-09-759-508B-1	Sequence 1, Appli	c 915	33	61.1	436	9	US-09-824-787B-24	Sequence 24, Appl
843	34	63.0	203654	10	US-09-820-908-3	Sequence 3, Appli	c 916	33	61.1	436	10	US-09-960-352-736	Sequence 736, App
844	34	63.0	368004	10	US-09-949-654-3	Sequence 3, Appli	c 917	33	61.1	436	10	US-09-960-352-4287	Sequence 4287, App
845	34	63.0	402850	9	US-09-844-653-5	Sequence 5, Appli	c 918	33	61.1	439	10	US-09-960-352-9006	Sequence 9006, App
846	34	63.0	684973	10	US-09-263-959-1	Sequence 1, Appli	c 919	33	61.1	442	9	US-10-007-280A-132	Sequence 132, App
847	33.5	62.0	258	10	US-09-974-300-4000	Sequence 4000, Ap	c 920	33	61.1	443	10	US-09-833-381-160	Sequence 160, App
848	33.5	62.0	73467	9	US-10-237-859-3	Sequence 3, Appli	c 921	33	61.1	446	9	US-09-918-995-14325	Sequence 14325, A
849	33	61.1	100	10	US-09-728-446-860	Sequence 860, App	c 922	33	61.1	446	10	US-09-960-352-8800	Sequence 8800, App
850	33	61.1	110	10	US-09-783-590-9713	Sequence 9713, Ap	c 923	33	61.1	447	9	US-09-918-995-23962	Sequence 23962, A
851	33	61.1	129	10	US-09-864-761-18241	Sequence 18241, A	c 924	33	61.1	448	9	US-09-918-995-27627	Sequence 27627, A
852	33	61.1	169	10	US-09-864-761-27083	Sequence 27083, A	c 925	33	61.1	449	9	US-09-918-995-28133	Sequence 28133, A
853	33	61.1	177	10	US-09-864-761-26695	Sequence 26695, A	c 926	33	61.1	450	9	US-09-918-995-21549	Sequence 21549, A
854	33	61.1	198	10	US-09-864-761-27521	Sequence 27521, A	c 927	33	61.1	453	9	US-09-918-995-19398	Sequence 19398, A
855	33	61.1	224	10	US-09-244-694-64	Sequence 64, Appl	c 928	33	61.1	453	10	US-09-864-761-26951	Sequence 26951, A
856	33	61.1	227	10	US-09-960-352-14739	Sequence 14739, A	c 929	33	61.1	453	10	US-09-880-107-1085	Sequence 1085, App
857	33	61.1	240	9	US-09-938-842A-686	Sequence 686, App	c 930	33	61.1	461	9	US-09-918-995-12884	Sequence 12884, A
858	33	61.1	245	9	US-09-764-891-9795	Sequence 9795, Ap	c 931	33	61.1	462	9	US-09-918-995-12732	Sequence 12732, A
859	33	61.1	250	10	US-09-974-300-8066	Sequence 8066, Ap	c 932	33	61.1	462	9	US-09-918-995-13690	Sequence 13690, A
860	33	61.1	267	9	US-09-736-457-1165	Sequence 1165, Ap	c 933	33	61.1	462	10	US-09-960-352-10152	Sequence 10152, A
861	33	61.1	267	9	US-09-902-941-1165	Sequence 1165, Ap	c 934	33	61.1	465	10	US-09-867-701-6197	Sequence 6197, App
862	33	61.1	267	9	US-09-849-626-1165	Sequence 1165, Ap	c 935	33	61.1	466	10	US-09-864-761-1281	Sequence 1281, App
863	33	61.1	267	9	US-10-017-754-1165	Sequence 1165, Ap	c 936	33	61.1	467	9	US-09-736-457-456	Sequence 456, App
864	33	61.1	286	10	US-09-923-876-5538	Sequence 5538, Ap	c 937	33	61.1	467	9	US-09-902-941-456	Sequence 456, App
865	33	61.1	297	10	US-09-783-590-1345	Sequence 1345, Ap	c 938	33	61.1	467	9	US-09-796-692-7784	Sequence 7784, App
866	33	61.1	314	9	US-09-764-891-1082	Sequence 1082, Ap	c 939	33	61.1	467	9	US-09-796-692-8149	Sequence 8149, App
867	33	61.1	314	10	US-09-864-761-32286	Sequence 32286, A	c 940	33	61.1	467	9	US-09-796-692-8442	Sequence 8442, App
868	33	61.1	322	10	US-09-983-965-5711	Sequence 5711, Ap	c 941	33	61.1	467	9	US-09-849-626-456	Sequence 456, App
869	33	61.1	328	10	US-09-960-352-10281	Sequence 10281, A	c 942	33	61.1	467	9	US-10-017-754-456	Sequence 456, App
870	33	61.1	330	9	US-09-918-995-24107	Sequence 24107, A	c 943	33	61.1	467	9	US-10-040-862-7784	Sequence 7784, App
871	33	61.1	340	10	US-09-983-965-1036	Sequence 1036, Ap	c 944	33	61.1	467	9	US-10-040-862-8149	Sequence 8149, App
872	33	61.1	342	10	US-09-983-965-3610	Sequence 3610, Ap	c 945	33	61.1	467	9	US-10-040-862-8442	Sequence 8442, App
873	33	61.1	351	10	US-09-864-761-18043	Sequence 18043, A	c 946	33	61.1	468	9	US-09-918-995-37274	Sequence 37274, A
874	33	61.1	359	9	US-09-803-719-1747	Sequence 1747, Ap	c 947	33	61.1	469	10	US-09-864-761-10053	Sequence 10053, A
875	33	61.1	360	9	US-09-918-995-37700	Sequence 37700, A	c 948	33	61.1	469	10	US-09-864-761-15780	Sequence 15780, A
876	33	61.1	360	10	US-09-960-352-3127	Sequence 3127, Ap	c 949	33	61.1	472	10	US-09-770-444-143	Sequence 143, App
877	33	61.1	366	9	US-09-822-846-388	Sequence 388, App	c 950	33	61.1	473	10	US-09-867-701-3683	Sequence 3683, App
878	33	61.1	368	9	US-10-046-935-1248	Sequence 1248, Ap	c 951	33	61.1	474	9	US-09-918-995-30884	Sequence 30884, A
879	33	61.1	368	9	US-09-878-178-1248	Sequence 1248, Ap	c 952	33	61.1	474	9	US-09-918-995-32246	Sequence 32246, A
880	33	61.1	368	9	US-10-146-502-1248	Sequence 1248, Ap	c 953	33	61.1	477	9	US-09-918-995-19651	Sequence 19651, A

```
c 954 33 61.1 479 9 US-09-918-995-947
c 955 33 61.1 480 10 US-09-864-761-10882
c 956 33 61.1 487 9 US-09-918-995-12690
c 957 33 61.1 487 9 US-09-918-995-29221
c 958 33 61.1 488 9 US-09-918-995-19661
c 959 33 61.1 490 9 US-09-918-995-31321
c 960 33 61.1 491 10 US-09-783-590-8388
c 961 33 61.1 496 9 US-09-918-995-31443
c 962 33 61.1 497 10 US-09-783-590-5698
c 963 33 61.1 501 9 US-09-918-995-21595
c 964 33 61.1 504 9 US-09-796-692-8076
c 965 33 61.1 504 9 US-10-040-862-8076
c 966 33 61.1 505 10 US-09-783-590-11224
c 967 33 61.1 505 10 US-09-867-701-10877
c 968 33 61.1 510 9 US-09-918-995-24236
c 969 33 61.1 510 10 US-09-867-701-3222
c 970 33 61.1 513 9 US-09-918-995-22037
c 971 33 61.1 517 9 US-09-764-891-9575
c 972 33 61.1 527 9 US-09-764-891-9576
c 973 33 61.1 537 9 US-09-764-891-2338
c 974 33 61.1 538 9 US-10-109-548-1
c 975 33 61.1 544 9 US-09-918-995-2207
c 976 33 61.1 549 9 US-10-027-806-27
c 977 33 61.1 549 9 US-10-034-623-27
c 978 33 61.1 549 9 US-10-027-801-27
c 979 33 61.1 559 9 US-10-060-036-1381
c 980 33 61.1 570 10 US-09-918-686-15
c 981 33 61.1 575 10 US-09-864-761-9210
c 982 33 61.1 578 10 US-09-864-761-1532
c 983 33 61.1 583 10 US-09-864-761-13527
c 984 33 61.1 585 12 US-10-001-843-62
c 985 33 61.1 606 10 US-09-770-149-914
c 986 33 61.1 607 10 US-09-864-761-33001
c 987 33 61.1 609 10 US-09-815-242-7744
c 988 33 61.1 620 9 US-09-828-995B-54
c 989 33 61.1 620 9 US-09-828-995B-56
c 990 33 61.1 704 9 US-09-892-877-103
c 991 33 61.1 719 9 US-09-764-868-448
c 992 33 61.1 719 9 US-09-955-999-17
c 993 33 61.1 731 10 US-09-770-149-102
c 994 33 61.1 732 10 US-09-974-300-6813
c 995 33 61.1 734 9 US-10-002-050-3
c 996 33 61.1 734 9 US-10-002-304-3
c 997 33 61.1 734 12 US-10-003-152-3
c 998 33 61.1 735 9 US-09-764-891-8245
c 999 33 61.1 772 10 US-09-764-853-175
1000 33 61.1 790 10 US-09-770-445-848

ALIGNMENTS

RESULT 1
US-08-781-986A-752
; Sequence 752, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A

Sequence 947, App
Sequence 10882, A
Sequence 12690, A
Sequence 29221, A
Sequence 19661, A
Sequence 31321, A
Sequence 8388, Ap
Sequence 31443, A
Sequence 5698, Ap
Sequence 21695, A
Sequence 8076, Ap
Sequence 8076, Ap
Sequence 11224, A
Sequence 10877, A
Sequence 24236, A
Sequence 3222, Ap
Sequence 22037, A
Sequence 9575, Ap
Sequence 9576, Ap
Sequence 2338, Ap
Sequence 1, Appli
Sequence 2207, Ap
Sequence 27, Appl
Sequence 27, Appl
Sequence 1381, Ap
Sequence 15, Appl
Sequence 9210, Ap
Sequence 1532, A
Sequence 13527, A
Sequence 62, Appl
Sequence 914, App
Sequence 33001, A
Sequence 7744, Ap
Sequence 54, Appl
Sequence 56, Appl
Sequence 103, App
Sequence 448, App
Sequence 17, Appl
Sequence 102, App
Sequence 6813, Ap
Sequence 3, Appli
Sequence 3, Appli
Sequence 8245, Ap
Sequence 175, App
Sequence 848, App

US-10-008-355-25 (1-10) x US-08-781-986A-752 (1-1019)
Pred. No.: 0.316 Length: 1019
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
|||||
Db 642 ACTGGTGGTAATTCAGGTTCACCTGTATT 671

RESULT 2
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-1
Alignment Scores:
Pred. No.: 0.655 Length: 2139
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-008-355-1 (1-2139)
QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
|||||
Db 1930 ACGGGCGGTAACCTCCGGTAGCCCGGTATTC 1959

RESULT 3
US-09-918-995-37867
; Sequence 37867, Application US/09918995
```

```

US-10-008-355-25 (1-10) x US-09-764-891-18 (1-423)

Qy 1 ThrGlyGlyAsnSerGlySerProValphe 10
   |||||  ::|||  ::|||  ::|||
Db 300 ACGGCGGAGAAATGCGAGTCATCTTC 329

RESULT 5
US-09-070-927A-261/c
; Sequence 261, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;           Patrick J. Dillon
;           Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 261:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4951 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 261:
;
US-09-070-927A-261

Alignment Scores:
Pred. No.: 101 Length: 4951
Score: 44.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 81.48% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-070-927A-261 (1-4951)

Qy 1 ThrGlyGlyAsnSerGlySerProValphe 10
   |||||  |||||  |||||  |||||
Db 695 ACGGCGGTCAATCTGTTACCAATCTAT 666

RESULT 6
US-09-974-300-327
; Sequence 327, Application US/09974300
; Patent No. US20020146721A1

```

```
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-300-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 327
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-327

Alignment Scores:
Pred. No.: 17.2 Length: 534
Score: 43.00 Matches: 8
Percent Similarity: 88.89% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 1
Query Match: 79.63% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-974-300-327 (1-534)

QY 2 GlycylAsnSerGlySerProValPhe 10
||||| |||||||||
Db 357 GGAGGACAAAGCGGTTCACCGGTATTC 383

RESULT 7
US-10-070-676-5
; Sequence 5, Application US/10070676
; Publication No. US20030059788A1
; GENERAL INFORMATION:
; APPLICANT: Toque, Bruno
; APPLICANT: Bracco, Laurent
; APPLICANT: Schweighoffer, Fabien
; TITLE OF INVENTION: Genetic Markers of Toxicity, Preparation
; TITLE OF INVENTION: and Uses
; FILE REFERENCE: 50146/003002
; CURRENT APPLICATION NUMBER: US/10/070,676
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: PCT/FR00/02503
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: FR 99/11405
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/456,370
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 152
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-070-676-5

Alignment Scores:
Pred. No.: 7.62 Length: 152
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 77.78% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-070-676-5 (1-152)

QY 2 GlycylAsnSerGlySerProValPhe 10
||||| |||||||||
```

```
Db 26 GGGGGAACTCAGGCAATCCGATTTC 52

RESULT 8
US-09-918-995-31306
; Sequence 31306, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31306
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(471)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31306

Alignment Scores:
Pred. No.: 23.2 Length: 471
Score: 42.00 Matches: 6
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 0
Query Match: 77.78% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-09-918-995-31306 (1-471)

QY 2 GlycylAsnSerGlySerProValPhe 10
||||| |||||||||
Db 118 GGGGGAACTCAGGCAATCCGATTTC 144

RESULT 9
US-09-867-701-1705
; Sequence 1705, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1705
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(450)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-1705

Alignment Scores:
Pred. No.: 51.4 Length: 450
Score: 40.00 Matches: 7
Percent Similarity: 87.50% Conservative: 0
Best Local Similarity: 87.50% Mismatches: 1
Query Match: 74.07% Indels: 0
DB: 10 Gaps: 0
```

```

US-10-008-355-25 (1-10) x US-10-079-854-243 (1-7441)

Qy      2  GlyGlyAsnSerGlySerProVal 9
|||||.....:|||||
Db      2023  GCGGTAACACAGCAGCCCTGTG 2000

RESULT 12
US-09-764-878-243/c
; Sequence 243, Application US/09764878
; Patent No. US2002090615A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764.878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver.. 2.0
; SEQ ID NO 243
; LENGTH: 7441
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-878-243

Alignment Scores:
Pred. No.:      810      Length:      7441
Score:          40.00    Matches:      7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match:      74.07% Indels:      0
DB:              10      Gaps:      0

US-10-008-355-25 (1-10) x US-09-764-878-243 (1-7441)

Qy      2  GlyGlyAsnSerGlySerProVal 9
|||||.....:|||||
Db      2023  GCGGTAACACAGCAGCCCTGTG 2000

RESULT 13
US-09-960-352-5672/c
; Sequence 5672, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5672
; LENGTH: 235
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 24-LIB34-067-Q1-E1-F7
US-09-960-352-5672

Alignment Scores:
Pred. No.:      41.4      Length:      235
Score:          39.00    Matches:      7
Percent Similarity: 77.78% Conservative: 0
Best Local Similarity: 77.78% Mismatches: 2
Query Match:      72.22% Indels:      0
DB:              10      Gaps:      0

US-10-008-355-25 (1-10) x US-09-960-352-5672 (1-235)

Qy      2  GlyGlyAsnSerGlySerProValPhe 10
|||||.....:|||||

```

```
Db 211 GGAGGCAGGTCAGGTGGTCCAGTATTC 185
RESULT 14
US-09-918-995-22868
; Sequence 22868, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22868
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(479)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22868
Alignment Scores:
Pred. No.: 83.3 Length: 479
Score: 39.00 Matches: 7
Percent Similarity: 77.78% Conservative: 0
Best Local Similarity: 77.78% Mismatches: 2
Query Match: 72.22% Indels: 0
Gaps: 0
DB: 9
US-10-008-355-25 (1-10) x US-09-918-995-22868 (1-479)
Qy 2 GlycAsnSerGlySerProValPhe 10
||||||| ||| |||||||
Db 58 GGGGGAACGTGGGGAACCTGTGTTT 84
RESULT 15
US-09-864-761-9954/c
; Sequence 9954, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomlca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9954
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007463.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
US-09-864-761-9954
Alignment Scores:
Pred. No.: 101 Length: 584
Score: 39.00 Matches: 6
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 72.22% Indels: 0
Gaps: 0
DB: 10
US-10-008-355-25 (1-10) x US-09-864-761-9954 (1-584)
Qy 3 GlyAsnSerGlySerProValPhe 10
||||||| ||| |||||||
Db 160 GGGAACTCTGGACCCCATATTT 137
RESULT 16
US-09-893-519A-136/c
; Sequence 136, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: ZHU, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/IG548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
```

; NUMBER OF SEQ ID NOS: 146  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 136  
; LENGTH: 1173  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-893-519A-136

Alignment Scores:  
Pred. No.: 201 Length: 1173  
Score: 39.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-09-893-519A-136 (1-1173)

Qy 2 GlyGlyAsnSerGlySerPro 8  
Db 101 GGAGGTAATTCAGGTACCT 81

## RESULT 17

US-10-001-887-75  
; Sequence 75, Application US/10001887  
; Patent No.: US200201534641  
; GENERAL INFORMATION:  
; APPLICANT: Salceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Caferkey, Robert  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Proteins  
; FILE REFERENCE: DEX-0269  
; CURRENT APPLICATION NUMBER: US/10/001,887  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: 60/249,998  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: 60/252,563  
; PRIOR FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 75  
; LENGTH: 1775  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-001-887-75

Alignment Scores:  
Pred. No.: 302 Length: 1775  
Score: 39.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-001-887-75 (1-1775)

Qy 2 GlyGlyAsnSerGlySerPro 8  
Db 396 GGGGGGAATTCAGGTACCA 416

## RESULT 18

US-09-764-868-111  
; Sequence 111, Application US/09764868  
; Patent No.: US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT32  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 111  
; LENGTH: 2109  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-868-111

Alignment Scores:  
Pred. No.: 358 Length: 2109  
Score: 39.00 Matches: 7  
Percent Similarity: 77.78% Conservative: 0  
Best Local Similarity: 77.78% Mismatches: 2  
Query Match: 72.22% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-09-764-868-111 (1-2109)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10  
Db 1809 GGGGGGAACGTGGGGAACTGTGTTT 1835

## RESULT 19

US-10-025-380-683  
; Sequence 683, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yudi  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick Thomas S.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; FILE REFERENCE: 210121.471C14  
; CURRENT APPLICATION NUMBER: US/10/025,380  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 683  
; LENGTH: 3255  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-025-380-683

Alignment Scores:  
Pred. No.: 548 Length: 3255  
Score: 39.00 Matches: 7  
Percent Similarity: 77.78% Conservative: 0  
Best Local Similarity: 77.78% Mismatches: 2  
Query Match: 72.22% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-025-380-683 (1-3255)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10  
Db 2988 GGGGGGAACGTGGGGAACTGTGTTT 3014

## RESULT 20

US-09-922-217-683

US-10-008-355-25 (1-10) x US-09-833-263-683 (1-3255)

RESULT 24



```
US-09-815-343-676
; Sequence 676, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 676
; LENGTH: 222
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-676

Alignment Scores:
Pred. No.: 59.6      Length: 222
Score: 38.00      Matches: 7
Percent Similarity: 90.00%      Conservative: 2
Best Local Similarity: 70.00%      Mismatches: 1
Query Match: 70.37%      Indels: 0
DB: 10      Gaps: 0

US-10-008-355-25 (1-10) x US-09-815-343-676 (1-222)
QY 1 ThrGlycylAsnSerGlySerProValPhe 10
||| |||:::|||||
Db 96 ACTCAGCGCAGTTCGTGGATCACCTGTGTAT 125

RESULT 25
US-09-815-343-1285
; Sequence 1285, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1285
; LENGTH: 222
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-1285

Alignment Scores:
Pred. No.: 59.6      Length: 222
Score: 38.00      Matches: 7
Percent Similarity: 90.00%      Conservative: 2
Best Local Similarity: 70.00%      Mismatches: 1
Query Match: 70.37%      Indels: 0
DB: 10      Gaps: 0

US-10-008-355-25 (1-10) x US-09-815-343-1285 (1-222)
QY 1 ThrGlycylAsnSerGlySerProValPhe 10
||| |||:::|||||
Db 96 ACTCAGCGCAGTTCGTGGATCACCTGTGTAT 125

RESULT 26
US-09-815-343-614
; Sequence 614, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 614
; LENGTH: 243
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-614

Alignment Scores:
Pred. No.: 65.1      Length: 243
Score: 38.00      Matches: 7
Percent Similarity: 90.00%      Conservative: 2
Best Local Similarity: 70.00%      Mismatches: 1
Query Match: 70.37%      Indels: 0
DB: 10      Gaps: 0

US-10-008-355-25 (1-10) x US-09-815-343-614 (1-243)
QY 1 ThrGlycylAsnSerGlySerProValPhe 10
||| |||:::|||||
Db 96 ACTCAGCGCAGTTCGTGGATCACCTGTGTAT 125

RESULT 27
US-09-815-343-729
; Sequence 729, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 729
; LENGTH: 253
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-729

Alignment Scores:
Pred. No.: 67.8      Length: 253
Score: 38.00      Matches: 7
Percent Similarity: 90.00%      Conservative: 2
Best Local Similarity: 70.00%      Mismatches: 1
Query Match: 70.37%      Indels: 0
DB: 10      Gaps: 0

US-10-008-355-25 (1-10) x US-09-815-343-729 (1-253)
QY 1 ThrGlycylAsnSerGlySerProValPhe 10
||| |||:::|||||
Db 96 ACTCAGCGCAGTTCGTGGATCACCTGTGTAT 125
```

```
RESULT 28
US-09-796-692-5591/c
: Sequence 5591, Application US/09796692
: Publication No. US20020198362A1
: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Algate, Paul A.
: APPLICANT: Mannion, Jane
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
: FILE REFERENCE: 2077.001200
: CURRENT APPLICATION NUMBER: US/09/796,692
: CURRENT FILING DATE: 2001-03-01
: PRIOR APPLICATION NUMBER: 60/186,126
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: 60/190,479
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: 60/200,545
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 60/200,303
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,779
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,999
: PRIOR FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 60/186,126
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: 60/190,479
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: 60/200,545
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 60/200,303
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,779
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,999
: PRIOR FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: 60/202,084
: PRIOR FILING DATE: 2000-05-04
: PRIOR APPLICATION NUMBER: 60/206,201
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: 60/218,950
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: 60/222,903
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: 60/223,416
: PRIOR FILING DATE: 2000-08-04
: PRIOR APPLICATION NUMBER: 60/223,378
: PRIOR FILING DATE: 2000-08-07
: PRIOR APPLICATION NUMBER: 60/202,084
: PRIOR FILING DATE: 2000-05-04
: PRIOR APPLICATION NUMBER: 60/206,201
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: 60/218,950
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: 60/222,903
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: 60/223,416
: PRIOR FILING DATE: 2000-08-04
: PRIOR APPLICATION NUMBER: 60/223,378
: PRIOR FILING DATE: 2000-08-07
: NUMBER OF SEQ ID NOS: 9597
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5591
: LENGTH: 319
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-796-692-5591

Alignment Scores:
Pred. No.: 85.1 Length: 319
Score: 38.00 Matches: 6
Percent Similarity: 90.00% Conservative: 3
Best Local Similarity: 60.00% Mismatches: 1
Query Match: 70.37% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-09-796-692-5591 (1-319)
Qy 1 ThrGlyGlyAsnSerGlySerProValPhe 10
Db 44 TCTGGAGGCTCTCAGGGGACCCCTGTCTTC 15

RESULT 29
US-10-040-862-5591/c
: Sequence 5591, Application US/10040862
: Publication No. US20030078396A1
: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Algate, Paul A.
: APPLICANT: Mannion, Jane
: APPLICANT: Retter, Marc
: APPLICANT: Corixa Corporation
: TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
: FILE REFERENCE: 014058-01352005
```

```
: CURRENT APPLICATION NUMBER: US/10/040,862
: CURRENT FILING DATE: 2001-11-06
: PRIOR APPLICATION NUMBER: US 60/186,126
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: US 60/190,479
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: US 60/200,545
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: US 60/200,303
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: US 60/200,779
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: US 60/200,999
: PRIOR FILING DATE: 2000-05-01
: PRIOR APPLICATION NUMBER: US 60/202,084
: PRIOR FILING DATE: 2000-05-04
: PRIOR APPLICATION NUMBER: US 60/206,201
: PRIOR FILING DATE: 2000-05-22
: PRIOR APPLICATION NUMBER: US 60/218,950
: PRIOR FILING DATE: 2000-07-14
: PRIOR APPLICATION NUMBER: US 60/222,903
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: US 60/223,416
: PRIOR FILING DATE: 2000-08-04
: PRIOR APPLICATION NUMBER: US 60/223,378
: PRIOR FILING DATE: 2000-08-07
: PRIOR APPLICATION NUMBER: US 09/796,692
: PRIOR FILING DATE: 2001-03-01
: NUMBER OF SEQ ID NOS: 10467
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5591
: LENGTH: 319
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-040-862-5591

Alignment Scores:
Pred. No.: 85.1 Length: 319
Score: 38.00 Matches: 6
Percent Similarity: 90.00% Conservative: 3
Best Local Similarity: 60.00% Mismatches: 1
Query Match: 70.37% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-040-862-5591 (1-319)
Qy 1 ThrGlyGlyAsnSerGlySerProValPhe 10
Db 44 TCTGGAGGCTCTCAGGGGACCCCTGTCTTC 15

RESULT 30
US-09-796-692-6270
: Sequence 6270, Application US/09796692
: Publication No. US20020198362A1
: GENERAL INFORMATION:
: APPLICANT: Gaiger, Alexander
: APPLICANT: Algate, Paul A.
: APPLICANT: Mannion, Jane
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
: FILE REFERENCE: 2077.001200
: CURRENT APPLICATION NUMBER: US/09/796,692
: CURRENT FILING DATE: 2001-03-01
: PRIOR APPLICATION NUMBER: 60/186,126
: PRIOR FILING DATE: 2000-03-01
: PRIOR APPLICATION NUMBER: 60/190,479
: PRIOR FILING DATE: 2000-03-17
: PRIOR APPLICATION NUMBER: 60/200,545
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 60/200,303
: PRIOR FILING DATE: 2000-04-28
: PRIOR APPLICATION NUMBER: 60/200,779
: PRIOR FILING DATE: 2000-04-28
```

```

/ PRIOR FILING DATE: 2000-05-22
/ PRIOR APPLICATION NUMBER: US 60/218,950
/ PRIOR FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: US 60/222,903
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: US 60/223,416
/ PRIOR FILING DATE: 2000-08-04
/ PRIOR APPLICATION NUMBER: US 60/223,378
/ PRIOR FILING DATE: 2000-08-07
/ PRIOR APPLICATION NUMBER: US 09/796,692
/ PRIOR FILING DATE: 2001-03-01
/ NUMBER OF SEQ ID NOS: 10467
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 6270
/ LENGTH: 320
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)
/ OTHER INFORMATION: n=A,T,C or G
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (43)
/ OTHER INFORMATION: n=A,T,C or G
US-10-040-862-6270

Alignment Scores:
Pred. No.: 85.4 Length: 320
Score: 38.00 Matches: 6
Percent Similarity: 90.00% Conservative: 3
Best Local Similarity: 60.00% Mismatches: 1
Query Match: 70.37% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-040-862-6270 (1-320)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
:::|||||::: |||:::|||||||
Db 277 TCTGGAGGCTCTCAGGGACCCCTGCTTC 306

RESULT 32
US-09-770-444-591
/ Sequence 591, Application US/09770444
/ Patent No. US20020023280A1
/ GENERAL INFORMATION:
/ APPLICANT: Gorlach, Jorn
/ APPLICANT: An, Yong-Qiang
/ APPLICANT: Hamilton, Carol M.
/ APPLICANT: Price, Jennifer L.
/ APPLICANT: Raines, Tracy M.
/ APPLICANT: Yu, Yang
/ APPLICANT: Rameaka, Joshua G.
/ APPLICANT: Page, Amy
/ APPLICANT: Matthew, Abraham V.
/ APPLICANT: Ledford, Brooke L.
/ APPLICANT: Woessner, Jeffrey P.
/ APPLICANT: Haas, William David
/ APPLICANT: Garcia, Carlos A.
/ APPLICANT: Kricker, Maja
/ APPLICANT: Slader, Ted
/ APPLICANT: Davis, Keith R.
/ APPLICANT: Allen, Keith
/ APPLICANT: Hoffman, Neil
/ APPLICANT: Hurban, Patrick
/ TITLE OF INVENTION: Expressed Sequences of Arabidopsis
/ TITLE OF INVENTION: thaliana
/ FILE REFERENCE: 2027 (PARA-016PRV)
/ CURRENT APPLICATION NUMBER: US/09/770,444
/ CURRENT FILING DATE: 2001-01-26
/ PRIOR APPLICATION NUMBER: 60/178,502
/ PRIOR FILING DATE: 2000-01-27
/ NUMBER OF SEQ ID NOS: 999

```

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 591
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-444-591
```

```
Alignment Scores:
Pred. No.: 120 Length: 454
Score: 38.00 Matches: 7
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 87.50% Mismatches: 0
Query Match: 70.37% Indels: 0
DB: 10 Gaps: 0
```

US-10-008-355-25 (1-10) x US-09-770-444-591 (1-454)

```
QY 2 GlyClyAsnSerGlySerProVal 9
|||||:|||||
DB 202 GGTGGTTCAAGTGTCTCCCGTC 225
```

RESULT 33

```
US-09-918-995-340
; Sequence 340, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 460
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(460)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-340
```

```
Alignment Scores:
Pred. No.: 122 Length: 460
Score: 38.00 Matches: 7
Percent Similarity: 87.50% Conservative: 0
Best Local Similarity: 87.50% Mismatches: 1
Query Match: 70.37% Indels: 0
DB: 9 Gaps: 0
```

US-10-008-355-25 (1-10) x US-09-918-995-340 (1-460)

```
QY 1 ThrGlyGlyAsnSerGlySerPro 8
||| |||||||
DB 307 ACTCTGGAAACTCTGGATCTCCT 330
```

RESULT 34

```
US-09-867-701-5230/c
; Sequence 5230, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
```

```
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5230
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-5230
```

```
Alignment Scores:
Pred. No.: 126 Length: 474
Score: 38.00 Matches: 7
Percent Similarity: 70.00% Conservative: 0
Best Local Similarity: 70.00% Mismatches: 3
Query Match: 70.37% Indels: 0
DB: 10 Gaps: 0
```

US-10-008-355-25 (1-10) x US-09-867-701-5230 (1-474)

```
QY 1 ThrClyGlyAsnSerGlySerProValPhe 10
|||||:|||||
DB 402 ACCGGGGTCTCTCTGCTCCCGTACTTC 373
```

RESULT 35

```
US-09-783-590-6501/c
; Sequence 6501, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR FILING DATE: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6501
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (110)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (356)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (397)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (413)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (421)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (458)
```

OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (463)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (464)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-6501

Alignment Scores:  
Pred. No.: 130 Length: 491  
Score: 38.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 87.50% Mismatches: 0  
Query Match: 70.37% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-783-590-6501 (1-491)

QY 2 GlyGlyAsnSerGlySerProVal 9  
|||||:|||||

Db 215 GGTGGTCAAGTGGTTCACCCGTG 192

RESULT 36

US-09-783-590-4015

; Sequence 4015, Application US/09783590

; Patent No. US20020110850A1

; GENERAL INFORMATION:

; APPLICANT: Dillion, Patrick J.

; APPLICANT: Haseltine, William A.

; APPLICANT: Li, Haodong

; APPLICANT: Rosen, Craig A.

; APPLICANT: Ruben, Steven M.

; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2

; FILE REFERENCE: PO-16.2C1

; CURRENT APPLICATION NUMBER: US/09/783,590

; CURRENT FILING DATE: 2000-02-15

; PRIOR APPLICATION NUMBER: 08/420,856

; PRIOR FILING DATE: 1995-04-12

; PRIOR APPLICATION NUMBER: 08/346,731

; PRIOR FILING DATE: 1994-11-21

; NUMBER OF SEQ ID NOS: 12485

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4015

; LENGTH: 508

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (10)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (315)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (359)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (366)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (372)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (381)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (394)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (398)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

LOCATION: (404)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (424)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (432)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (448)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (454)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (468)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (489)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-4015

Alignment Scores:

Pred. No.: 134 Length: 508  
Score: 38.00 Matches: 6  
Percent Similarity: 80.00% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 2  
Query Match: 70.37% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-783-590-4015 (1-508)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10

Db 456 ACTGGGGCAATNGGGGGCCCAATTTT 485

RESULT 37

US-10-076-622-123

; Sequence 123, Application US/10076622

; Publication No. US20030023036A1

; GENERAL INFORMATION:

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: Persing, David H.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.470C11

; CURRENT APPLICATION NUMBER: US/10/076,622

; CURRENT FILING DATE: 2002-02-13

; NUMBER OF SEQ ID NOS: 627

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 123

; LENGTH: 531

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: 152, 373, 482, 494, 496, 502

; OTHER INFORMATION: n = A,T,C or G

US-10-076-622-123

Alignment Scores:

Pred. No.: 140 Length: 531  
Score: 38.00 Matches: 6  
Percent Similarity: 77.78% Conservative: 1  
Best Local Similarity: 66.67% Mismatches: 2  
Query Match: 70.37% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-076-622-123 (1-531)

QY 2 GlyGlyAsnSerGlySerProValPhe 10

|||||

```
Db 241 GGGGAGGTGGGGAGCCCTATTTT 267
RESULT 38
US-09-604-287A-123
; Sequence 123, Application US/09604287A
; Patent No. US20020064872A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 123
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(531)
; OTHER INFORMATION: n = A,T,C or G
US-09-604-287A-123
Alignment Scores:
Pred. No.: 140 Length: 531
Score: 38.00 Matches: 6
Percent Similarity: 77.78% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 70.37% Indels: 0
Gaps: 0
DB: 10
US-10-008-355-25 (1-10) x US-09-604-287A-123 (1-531)
QY 2 GlycylAsnSerGlySerProValPhe 10
||||| |||||||:|||||
Db 241 GGGGAGGTGGGGAGCCCTATTTT 267
RESULT 40
US-10-007-805-123
; Sequence 123, Application US/10007805
; Patent No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 152_373, 482, 494, 496, 502
; OTHER INFORMATION: n = A,T,C or G
US-10-007-805-123
Alignment Scores:
Pred. No.: 140 Length: 531
Score: 38.00 Matches: 6
Percent Similarity: 77.78% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 70.37% Indels: 0
Gaps: 0
DB: 12
US-10-008-355-25 (1-10) x US-10-007-805-123 (1-531)
QY 2 GlycylAsnSerGlySerProValPhe 10
||||| |||||||:|||||
Db 241 GGGGAGGTGGGGAGCCCTATTTT 267
RESULT 41
US-09-918-995-27317/c
; Sequence 27317, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 27317
; LENGTH: 562
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)....(562)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27317

Alignment Scores:
Pred. No.: 148
Score: 38.00
Percent Similarity: 88.89%
Best Local Similarity: 77.78%
Query Match: 70.37%
Indels: 0
Gaps: 0

US-10-008-355-25 (1-10) x US-09-918-995-27317 (1-562)
QY 2 GlyGlyAsnSerGlySerProValPhe 10
|||||
Db 547 GGAGGTAACCTGGGATGACCCCGTGTAC 521

RESULT 42
US-09-808-701-14
; Sequence 14, Application US/09808701
; Patent No. US20020146757A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Kyle
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ren, Feiyan
; APPLICANT: Wang, Dunrui
; APPLICANT: Chen, Rui-hong
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020146757A1el Nucleic Acids and
; FILE REFERENCE: 790CIP2D
; CURRENT APPLICATION NUMBER: US/09/808,701
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 14
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (131)..(952)
US-09-808-701-14

Alignment Scores:
Pred. No.: 291
Score: 38.00
Percent Similarity: 77.78%
Best Local Similarity: 77.78%
Query Match: 70.37%
Indels: 0
Gaps: 0

US-10-008-355-25 (1-10) x US-09-808-701-14 (1-1114)
QY 1 ThrGlyGlyAsnSerGlySerProVal 9
|||||
Db 526 ACCGGGGGAACCGAGGCTCCCGCAGTA 552

RESULT 43
US-10-043-487-33
; Sequence 33, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LEGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypep
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 1571
; TYPE: DNA
; ORGANISM: Shigella Flexneri
US-10-043-487-33

Alignment Scores:
Pred. No.: 408
Score: 38.00
Percent Similarity: 87.50%
Best Local Similarity: 87.50%
Query Match: 70.37%
Indels: 0
Gaps: 0

US-10-008-355-25 (1-10) x US-10-043-487-33 (1-1571)
QY 1 ThrGlyGlyAsnSerGlySerPro 8
|||||
Db 834 ACTCTGGAAACTCTGGATCTCCT 857

RESULT 44
US-09-764-864-332
; Sequence 332, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 332
; LENGTH: 2290
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-332

Alignment Scores:
Pred. No.: 591
Score: 38.00
Percent Similarity: 90.00%
Best Local Similarity: 70.00%
Query Match: 70.37%
Indels: 0
Gaps: 0

US-10-008-355-25 (1-10) x US-09-764-864-332 (1-2290)
QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
|||||
Db 1184 ACTCAGGCGAGTCTCGATCACCTGTGTAT 1213

RESULT 45
US-09-899-651-8
; Sequence 8, Application US/09899651
```

1.



; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PTZ23  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 329  
; LENGTH: 3694  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-764-864-329

Alignment Scores:  
Pred. No.: 945 Length: 3694  
Score: 38.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 70.37% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-764-864-329 (1-3694)

Qy 1 ThrGlyGlyAsnSerglySerProValPhe 10

Db 2558 ACTCAGCGCAGTCTGGATCACCCTGTGTAT 2587

RESULT 49

US-09-960-352-14943

; Sequence 14943, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byattal, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 14943

; LENGTH: 400

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 64-LIB2809-023-Q1-E1-H12

US-09-960-352-14943

Alignment Scores:

Pred. No.: 131 Length: 400  
Score: 37.50 Matches: 9  
Percent Similarity: 90.00% Conservative: 0  
Best Local Similarity: 90.00% Mismatches: 0  
Query Match: 69.44% Indels: 1  
DB: 10 Gaps: 1

US-10-008-355-25 (1-10) x US-09-960-352-14943 (1-400)

Qy 1 ThrGlyGlyAsn---SerglySerProVal 9

Db 110 ACAGCGGTAACCTCTCAGGCTCCAGTG 139

RESULT 50

US-09-864-761-32393/c

; Sequence 32393, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 32393  
; LENGTH: 222  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO ALL121716.16  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9  
; OTHER INFORMATION: SWISSPROT HIT: O52325, EVALUE 3.70e+00  
; OTHER INFORMATION: NT HIT: AB029040.1, EVALUE 1.00e-122  
; OTHER INFORMATION: EST\_HUMAN HIT: BE002805.1, EVALUE 1.00e-122  
US-09-864-761-32393

Alignment Scores:  
Pred. No.: 90.8 Length: 222  
Score: 37.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 75.00% Mismatches: 0  
Query Match: 68.52% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-864-761-32393 (1-222)

Qy 2 GlyGlyAsnSerglySerProVal 9

Db 69 GGAGCACACTCAGGAGTCCTATC 46

Search completed: May 23, 2003, 14:59:36

us-10-008-355-25.p2n.rnpb

---

Sun May 25 15:40:37 2003

Job time : 114 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 13:20:58 ; Search time 42 Seconds  
(without alignments)  
73.018 Million cells

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TGGNSGSPVF 10

Scoring table:	BLOSUM62
	Xgapop 10.0 , Xgapext 0.5
	Ygapop 10.0 , Ygapext 0.5
	Fgapop 6.0 , Fgapext 7.0
	Delop 6.0 , Delxt 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Command Line parameters:

```

-MODEL=frames_p2n.model -DRV=xlh
-Q/cgcn2_1/USTPO.spool/US10008355/runat_16052003_110502_3111/app_query.fasta_l.199
-DB=Issued_Patents_NA -OFTM=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=1000 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=50
-MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=20000000000
-USER=US10008355_@CGN_l_1_32_runat_16052003_110502_3111 -NCPUP=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT=LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : Issued patents NA:\*

```

1: /cgn2_6/poddata/1/ina/5A_COMB.seq.*
2: /cgn2_6/poddata/1/ina/5B_COMB.seq.*
3: /cgn2_6/poddata/1/ina/6A_COMB.seq.*
4: /cgn2_6/poddata/1/ina/6B_COMB.seq.*
5: /cgn2_6/poddata/1/ina/PCRTUS_COMB.seq.*
6: /cgn2_6/poddata/1/ina/backfiles1.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	%			Length	DB	ID	Description
	Score	Match	Query				
1	54	100.0	1974	4	US-09-221-017B-726		Sequence 726, App
2	49	90.7	936	4	US-09-134-001C-892		Sequence 892, App
3	48	88.9	2384	4	US-09-221-017B-1045		Sequence 1045, App
4	44	81.5	758	4	US-09-071-035-427		Sequence 427, App
5	44	81.5	888	4	US-09-071-035-425		Sequence 425, App
6	43	79.6	1448	1	US-08-035-634-1		Sequence 1, Appli
C 7	40	74.1	743	4	US-08-998-416-856		Sequence 856, App
8	39	72.2	240	2	US-08-626-169-2		Sequence 2, Appli
9	39	72.2	240	3	US-09-164-907-2		Sequence 2, Appli
10	39	72.2	2774	1	US-07-723-002C-5		Sequence 5, Appli
11	38	70.4	531	4	US-09-222-575-123		Sequence 123, App
C 12	38	70.4	663	1	US-07-624-313-3		Sequence 3, Appli

86	35	64.8	4899	2	US-08-210-535-5	Sequence 5, Appli	C 159	34	63.0	3223	2	US-08-620-694A-9	Sequence 9, Appli
87	35	64.8	4899	6	5268270-1	Patent No. 5268270	C 160	34	63.0	3223	3	US-09-022-255-9	Sequence 9, Appli
88	35	64.8	6619	2	US-08-683-007A-3	Sequence 3, Appli	C 161	34	63.0	3223	3	US-09-022-696-9	Sequence 9, Appli
89	35	64.8	8257	4	US-09-484-970B-65	Sequence 65, Appli	C 162	34	63.0	3223	3	US-08-978-773-3	Sequence 3, Appli
90	35	64.8	10642	4	US-09-934-551-3	Sequence 3, Appli	C 163	34	63.0	3223	3	US-09-022-253-9	Sequence 9, Appli
91	35	64.8	12001	1	US-08-458-568A-11	Sequence 11, Appli	C 164	34	63.0	3223	3	US-09-022-260-9	Sequence 9, Appli
92	35	64.8	44377	2	US-08-804-227C-7	Sequence 7, Appli	C 165	34	63.0	3223	4	US-09-022-259-9	Sequence 9, Appli
93	35	64.8	44377	2	US-08-804-198-1	Sequence 1, Appli	C 166	34	63.0	3223	4	US-09-022-257-9	Sequence 9, Appli
94	35	64.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli	C 167	34	63.0	3241	4	US-09-434-288-11	Sequence 11, Appli
95	35	64.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli	C 168	34	63.0	3287	4	US-08-811-481-15	Sequence 15, Appli
96	34.5	63.9	10906	4	US-08-976-259-63	Sequence 63, Appli	C 169	34	63.0	3357	4	US-09-221-017B-287	Sequence 287, App
97	34	63.0	70	2	US-08-488-402A-129	Sequence 129, App	C 170	34	63.0	3410	4	US-09-020-956-110	Sequence 110, App
98	34	63.0	70	2	US-08-484-552A-129	Sequence 129, App	C 171	34	63.0	3410	4	US-09-030-607-110	Sequence 110, App
99	34	63.0	70	5	PCT-US96-09472-129	Sequence 129, App	C 172	34	63.0	3410	4	US-09-605-785-110	Sequence 110, App
100	34	63.0	192	1	US-08-423-383-7	Sequence 7, Appli	C 173	34	63.0	3410	4	US-09-439-313-110	Sequence 110, App
101	34	63.0	192	2	US-08-437-353A-7	Sequence 7, Appli	C 174	34	63.0	3410	4	US-09-352-616A-110	Sequence 110, App
102	34	63.0	195	3	US-09-071-710-6	Sequence 6, Appli	C 175	34	63.0	3410	4	US-09-602-877A-100	Sequence 100, App
103	34	63.0	195	4	US-09-525-397-6	Sequence 6, Appli	C 176	34	63.0	3410	4	US-09-232-149A-110	Sequence 110, App
104	34	63.0	276	1	US-08-181-492B-25	Sequence 25, Appli	C 177	34	63.0	3891	2	US-08-969-630-3	Sequence 3, Appli
105	34	63.0	276	1	US-08-423-383-8	Sequence 8, Appli	C 178	34	63.0	4034	4	US-09-605-785-704	Sequence 704, App
106	34	63.0	276	5	US-08-181-492B-26	Sequence 26, Appli	C 179	34	63.0	4136	4	US-09-103-875-2	Sequence 2, Appli
107	34	63.0	380	5	PCT-US95-00408-25	Sequence 25, Appli	C 180	34	63.0	4894	4	US-09-605-785-702	Sequence 702, App
108	34	63.0	380	2	US-08-423-383-9	Sequence 9, Appli	C 181	34	63.0	6976	4	US-09-605-785-705	Sequence 705, App
109	34	63.0	410	1	US-08-437-353A-9	Sequence 9, Appli	C 182	34	63.0	8056	3	US-09-136-605-14	Sequence 14, Appli
110	34	63.0	410	2	US-08-776-971-118	Sequence 8, Appli	C 183	34	63.0	8060	4	US-08-766-528-1	Sequence 1, Appli
111	34	63.0	507	4	US-08-811-481-25	Sequence 5, Appli	C 184	34	63.0	8082	1	US-08-306-691B-41	Sequence 41, Appli
112	34	63.0	711	4	US-09-276-531-114	Sequence 114, App	C 185	34	63.0	8082	1	US-08-187-785-1	Sequence 1, Appli
113	34	63.0	738	4	US-08-811-481-24	Sequence 24, Appli	C 186	34	63.0	8082	4	US-09-167-322-11	Sequence 11, Appli
114	34	63.0	769	4	US-08-776-971-118	Sequence 118, App	C 187	34	63.0	8082	5	PCT-US93-06251-28	Sequence 28, Appli
115	34	63.0	769	4	US-08-776-971-119	Sequence 119, App	C 188	34	63.0	8132	4	US-08-766-528-3	Sequence 3, Appli
116	34	63.0	932	4	US-08-811-481-25	Sequence 25, Appli	C 189	34	63.0	10952	1	US-08-602-036A-1	Sequence 1, Appli
117	34	63.0	932	4	US-08-811-481-26	Sequence 26, Appli	C 190	34	63.0	10952	2	US-08-502-374A-1	Sequence 1, Appli
118	34	63.0	999	4	US-08-811-481-27	Sequence 27, Appli	C 191	34	63.0	10952	2	US-08-642-407A-1	Sequence 20, Appli
119	34	63.0	1011	4	US-08-811-481-27	Sequence 27, Appli	C 192	34	63.0	12368	4	US-08-976-259-20	Sequence 60, Appli
120	34	63.0	1147	1	US-08-417-103-15	Sequence 15, App	C 193	34	63.0	15652	4	US-09-422-936-6	Sequence 55, Appli
121	34	63.0	1244	1	US-07-816-283-7	Sequence 7, Appli	C 194	34	63.0	21040	4	US-08-961-527-55	Sequence 1, Appli
122	34	63.0	1244	1	US-08-417-103-7	Sequence 7, Appli	C 195	34	63.0	31571	1	US-08-323-443B-1	Sequence 1, Appli
123	34	63.0	1249	3	US-08-985-950-1	Sequence 1, Appli	C 196	34	63.0	45546	4	US-09-146-053-6	Sequence 6, Appli
124	34	63.0	1320	4	US-09-370-838-22	Sequence 22, Appli	C 197	34	63.0	53526	3	US-08-658-136-2	Sequence 2, Appli
125	34	63.0	1328	4	US-09-370-838-24	Sequence 24, Appli	C 198	34	63.0	53526	3	US-08-658-136-1	Sequence 1, Appli
126	34	63.0	1331	4	US-09-370-838-27	Sequence 27, Appli	C 199	34	63.0	53577	3	US-08-813-817-3	Sequence 3, Appli
127	34	63.0	1333	4	US-09-370-838-28	Sequence 28, Appli	C 200	34	63.0	59065	4	US-09-978-197-3	Sequence 3, Appli
128	34	63.0	1351	1	US-07-816-283-5	Sequence 5, Appli	C 201	34	63.0	59065	4	US-09-800-960-3	Sequence 3, Appli
129	34	63.0	1351	1	US-08-417-103-5	Sequence 5, Appli	C 202	34	63.0	62804	4	US-09-103-840A-2	Sequence 2, Appli
130	34	63.0	1355	4	US-09-370-838-31	Sequence 31, Appli	C 203	34	63.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
131	34	63.0	1381	4	US-09-695-458-21	Sequence 21, Appli	C 204	33	61.1	25	5281520-14	Patent No. 5281520	
132	34	63.0	1545	2	US-08-660-963-9	Sequence 9, Appli	C 205	33	61.1	42	PCT-US93-08329-18	Sequence 18, Appli	
133	34	63.0	1758	4	US-09-370-838-25	Sequence 25, Appli	C 206	33	61.1	132	1	US-08-450-945-68	Sequence 68, Appli
134	34	63.0	1869	4	US-09-350-268-1	Sequence 1, Appli	C 207	33	61.1	132	4	US-08-976-161-68	Sequence 68, Appli
135	34	63.0	1869	4	US-09-558-679-1	Sequence 1, Appli	C 208	33	61.1	300	4	US-09-060-756-402	Sequence 402, App
136	34	63.0	1997	2	US-08-969-630-1	Sequence 1, Appli	C 209	33	61.1	315	4	US-09-060-756-127	Sequence 127, App
137	34	63.0	2038	4	US-09-008-271A-18	Sequence 18, Appli	C 210	33	61.1	318	4	US-09-060-756-171	Sequence 171, App
138	34	63.0	2087	4	US-09-257-894-9	Sequence 9, Appli	C 211	33	61.1	350	4	US-09-060-756-397	Sequence 397, App
139	34	63.0	2143	3	US-09-071-710-15	Sequence 15, Appli	C 212	33	61.1	357	4	US-09-060-756-137	Sequence 137, App
140	34	63.0	2143	3	US-09-525-397-15	Sequence 15, Appli	C 213	33	61.1	441	4	US-09-060-756-24	Sequence 24, Appli
141	34	63.0	2152	3	US-09-071-710-16	Sequence 16, Appli	C 214	33	61.1	462	1	US-08-450-945-66	Sequence 66, Appli
142	34	63.0	2152	4	US-09-525-397-16	Sequence 16, Appli	C 215	33	61.1	462	4	US-08-976-161-66	Sequence 66, Appli
143	34	63.0	2165	4	US-09-257-894-8	Sequence 8, Appli	C 216	33	61.1	475	4	US-09-004-838-5	Sequence 5, Appli
144	34	63.0	2200	2	US-08-819-825-1	Sequence 1, Appli	C 217	33	61.1	587	3	US-09-332-934-8	Sequence 8, Appli
145	34	63.0	2200	4	US-08-163-642-1	Sequence 1, Appli	C 218	33	61.1	668	4	US-08-998-416-150	Sequence 150, App
146	34	63.0	2367	4	US-09-056-556-201	Sequence 201, App	C 219	33	61.1	677	4	US-08-858-207A-167	Sequence 167, App
147	34	63.0	2367	4	US-09-072-596-196	Sequence 196, App	C 220	33	61.1	717	4	US-08-998-416-586	Sequence 586, App
148	34	63.0	2383	4	US-09-192-909-1	Sequence 1, Appli	C 221	33	61.1	759	3	US-08-995-960-14	Sequence 14, Appli
149	34	63.0	2414	6	5248599-1	Patent No. 5248599	C 222	33	61.1	777	3	US-09-332-934-10	Sequence 10, Appli
150	34	63.0	2470	1	US-07-795-859B-1	Sequence 1, Appli	C 223	33	61.1	803	4	US-08-998-416-352	Sequence 52, App
151	34	63.0	2470	1	US-08-457-616-1	Sequence 1, Appli	C 224	33	61.1	852	4	US-09-171-209-51	Sequence 51, Appli
152	34	63.0	2494	3	US-08-714-918-3	Sequence 3, Appli	C 225	33	61.1	904	4	US-09-171-209-24	Sequence 24, Appli
153	34	63.0	2494	4	US-09-265-315-3	Sequence 3, Appli	C 226	33	61.1	950	6	5177308-3	Patent No. 5177308
154	34	63.0	2494	4	US-09-265-315-3	Sequence 3, Appli	C 227	33	61.1	978	3	US-09-267-031-13	Sequence 13, Appli
155	34	63.0	2494	4	US-09-266-417-3	Sequence 3, Appli	C 228	33	61.1	984	3	US-08-459-046-1	Sequence 1, Appli
156	34	63.0	2665	4	US-09-257-894-1	Sequence 1, Appli	C 229	33	61.1	999	4	US-08-818-112-52	Sequence 52, Appli
157	34	63.0	2725	3	US-08-941-445A-14	Sequence 14, Appli	C 230	33	61.1	999	4	US-08-818-111-52	Sequence 52, Appli
158	34	63.0	2736	4	US-08-811-481-23	Sequence 23, Appli	C 231	33	61.1	999	4	US-09-056-556-52	Sequence 52, Appli
159	34	63.0	2904	4	US-09-605-785-703	Sequence 703, App	C 232	33	61.1	999	4	US-09-072-596-52	Sequence 52, Appli

c 232	33	61.1	1070	4	US-09-470-443-7	Sequence 7, Appli	c 305	33	61.1	3350	1	US-08-247-946A-2	Sequence 2, Appli
c 233	33	61.1	1222	4	US-09-484-970B-3	Sequence 3, Appli	c 306	33	61.1	3350	5	PCT-US95-06420-2	Sequence 2, Appli
c 234	33	61.1	1282	1	US-08-361-920-22	Sequence 22, Appl	c 307	33	61.1	3409	2	US-08-658-665-54	Sequence 54, Appl
c 235	33	61.1	1282	1	US-08-479-939-22	Sequence 22, Appl	c 308	33	61.1	3409	4	US-08-796-101-18	Sequence 18, Appl
c 236	33	61.1	1282	1	US-08-483-432-22	Sequence 22, Appl	c 309	33	61.1	3409	4	US-09-085-273-54	Sequence 54, Appl
c 237	33	61.1	1352	5	PCT-US92-02091-7	Sequence 7, Appli	c 310	33	61.1	3499	2	US-08-658-665-52	Sequence 52, Appl
c 238	33	61.1	1383	2	US-08-658-665-58	Sequence 58, Appl	c 311	33	61.1	3499	4	US-08-796-101-16	Sequence 16, Appl
c 239	33	61.1	1383	4	US-08-796-101-22	Sequence 22, Appl	c 312	33	61.1	3499	4	US-09-085-273-52	Sequence 52, Appl
c 240	33	61.1	1386	2	US-09-085-273-58	Sequence 53, Appl	c 313	33	61.1	3572	4	US-09-072-596-337	Sequence 337, App
c 241	33	61.1	1386	4	US-08-658-665-53	Sequence 53, Appl	c 314	33	61.1	3622	2	US-08-658-665-60	Sequence 60, Appl
c 242	33	61.1	1386	4	US-08-796-101-17	Sequence 17, Appl	c 315	33	61.1	3622	4	US-08-796-101-24	Sequence 24, Appl
c 243	33	61.1	1386	4	US-09-085-273-53	Sequence 53, Appl	c 316	33	61.1	3622	4	US-09-085-273-60	Sequence 60, Appl
c 244	33	61.1	1390	1	US-08-628-417-1	Sequence 1, Appli	c 317	33	61.1	3637	1	US-08-445-640-3	Sequence 3, Appli
c 245	33	61.1	1434	5	PCT-US93-11404-1	Sequence 1, Appli	c 318	33	61.1	3637	3	US-08-170-558-3	Sequence 3, Appli
c 246	33	61.1	1464	3	US-08-605-150A-3	Sequence 3, Appli	c 319	33	61.1	3637	3	US-08-447-314-3	Sequence 3, Appli
c 247	33	61.1	1472	1	US-08-333-565-1	Sequence 1, Appli	c 320	33	61.1	3637	3	US-08-445-461-3	Sequence 3, Appli
c 248	33	61.1	1472	2	US-08-661-479-1	Sequence 1, Appli	c 321	33	61.1	3678	4	US-07-852-132A-14	Sequence 14, Appl
c 249	33	61.1	1476	2	US-08-658-665-50	Sequence 50, Appl	c 322	33	61.1	3688	6	5248670-4	Patent No. 5248670
c 250	33	61.1	1476	4	US-08-796-101-14	Sequence 14, Appl	c 323	33	61.1	3962	1	US-08-336-343A-1	Sequence 1, Appli
c 251	33	61.1	1476	4	US-09-085-273-50	Sequence 50, Appl	c 324	33	61.1	4097	3	US-09-123-708-5	Sequence 5, Appli
c 252	33	61.1	1656	3	US-09-026-958-1	Sequence 1, Appli	c 325	33	61.1	4097	3	US-09-123-624-5	Sequence 5, Appli
c 253	33	61.1	1671	1	US-08-385-191A-5	Sequence 5, Appli	c 326	33	61.1	4517	4	US-09-140-804-9	Sequence 9, Appli
c 254	33	61.1	1671	4	US-08-472-402A-5	Sequence 5, Appli	c 327	33	61.1	5100	1	US-08-164-292B-15	Sequence 15, Appl
c 255	33	61.1	1740	4	US-08-796-101-50	Sequence 50, Appl	c 328	33	61.1	5100	1	US-08-164-292B-17	Sequence 17, Appl
c 256	33	61.1	1748	3	US-09-100-730-1	Sequence 1, Appli	c 329	33	61.1	5100	1	US-08-164-292B-19	Sequence 19, Appl
c 257	33	61.1	1788	4	US-08-303-861-1	Sequence 1, Appli	c 330	33	61.1	5100	1	US-08-164-292B-21	Sequence 21, Appl
c 258	33	61.1	1794	4	US-09-342-647-3	Sequence 3, Appli	c 331	33	61.1	5100	1	US-08-164-292B-23	Sequence 23, Appl
c 259	33	61.1	1866	3	US-08-909-742-1	Sequence 1, Appli	c 332	33	61.1	5100	1	US-08-164-292B-25	Sequence 25, Appl
c 260	33	61.1	1866	4	US-09-412-289-1	Sequence 1, Appli	c 333	33	61.1	5100	3	US-08-845-623-15	Sequence 15, Appl
c 261	33	61.1	1926	4	US-09-249-585A-2	Sequence 2, Appli	c 334	33	61.1	5100	3	US-08-845-623-17	Sequence 17, Appl
c 262	33	61.1	1962	4	US-09-072-596-349	Sequence 349, App	c 335	33	61.1	5100	3	US-08-845-623-19	Sequence 19, Appl
c 263	33	61.1	1971	4	US-09-347-878-23	Sequence 23, Appl	c 336	33	61.1	5100	3	US-08-845-623-21	Sequence 21, Appl
c 264	33	61.1	1975	2	US-08-658-665-51	Sequence 51, Appl	c 337	33	61.1	5100	3	US-08-845-623-23	Sequence 23, Appl
c 265	33	61.1	1975	4	US-08-796-101-15	Sequence 15, Appl	c 338	33	61.1	5100	3	US-08-845-623-25	Sequence 25, Appl
c 266	33	61.1	1975	4	US-09-085-273-51	Sequence 51, Appl	c 339	33	61.1	5100	3	US-08-815-927-15	Sequence 15, Appl
c 267	33	61.1	2000	4	US-09-555-889A-1	Sequence 1, Appli	c 340	33	61.1	5100	3	US-08-815-927-17	Sequence 17, Appl
c 268	33	61.1	2033	1	US-08-148-910-14	Sequence 14, Appl	c 341	33	61.1	5100	3	US-08-815-927-19	Sequence 19, Appl
c 269	33	61.1	2033	1	US-08-448-937A-14	Sequence 14, Appl	c 342	33	61.1	5100	3	US-08-815-927-21	Sequence 21, Appl
c 270	33	61.1	2057	1	US-08-450-945-57	Sequence 57, Appl	c 343	33	61.1	5100	3	US-08-815-927-23	Sequence 23, Appl
c 271	33	61.1	2057	4	US-08-976-161-57	Sequence 57, Appl	c 344	33	61.1	5100	3	US-08-815-927-25	Sequence 25, Appl
c 272	33	61.1	2061	1	US-08-382-184-1	Sequence 1, Appli	c 345	33	61.1	5100	4	US-09-103-330-15	Sequence 15, Appl
c 273	33	61.1	2061	2	US-08-641-356-1	Sequence 1, Appli	c 346	33	61.1	5100	4	US-09-103-330-17	Sequence 17, Appl
c 274	33	61.1	2061	4	US-09-132-528-1	Sequence 1, Appli	c 347	33	61.1	5100	4	US-09-103-330-19	Sequence 19, Appl
c 275	33	61.1	2061	4	US-08-875-494-1	Sequence 1, Appli	c 348	33	61.1	5100	4	US-09-103-330-21	Sequence 21, Appl
c 276	33	61.1	2061	4	US-09-599-366-1	Sequence 1, Appli	c 349	33	61.1	5100	4	US-09-103-330-23	Sequence 23, Appl
c 277	33	61.1	2061	4	US-08-875-494-1	Sequence 1, Appli	c 350	33	61.1	5100	4	US-09-103-330-25	Sequence 25, Appl
c 278	33	61.1	2119	4	US-09-240-639-7	Sequence 7, Appli	c 351	33	61.1	5100	4	US-09-435-242-15	Sequence 15, Appl
c 279	33	61.1	2187	4	US-09-318-448-1	Sequence 1, Appli	c 352	33	61.1	5100	4	US-09-435-242-17	Sequence 17, Appl
c 280	33	61.1	2187	4	US-09-127-219B-2	Sequence 2, Appli	c 353	33	61.1	5100	4	US-09-435-242-19	Sequence 19, Appl
c 281	33	61.1	2219	3	US-08-738-000-3	Sequence 3, Appli	c 354	33	61.1	5100	4	US-09-435-242-21	Sequence 21, Appl
c 282	33	61.1	2219	4	US-09-258-928-3	Sequence 3, Appli	c 355	33	61.1	5100	4	US-09-435-242-23	Sequence 23, Appl
c 283	33	61.1	2220	3	US-08-738-000-1	Sequence 1, Appli	c 356	33	61.1	5100	4	US-09-435-242-25	Sequence 25, Appl
c 284	33	61.1	2220	4	US-09-258-928-1	Sequence 1, Appli	c 357	33	61.1	5175	4	US-08-972-927-4	Sequence 4, Appli
c 285	33	61.1	2225	1	US-08-450-945-56	Sequence 56, Appl	c 358	33	61.1	5452	2	US-09-130-114-1	Sequence 1, Appli
c 286	33	61.1	2225	4	US-08-976-161-56	Sequence 56, Appl	c 359	33	61.1	5463	4	US-09-470-443-1	Sequence 1, Appli
c 287	33	61.1	2230	4	US-09-342-647-1	Sequence 1, Appli	c 360	33	61.1	5482	4	US-09-470-443-3	Sequence 3, Appli
c 288	33	61.1	2379	4	US-09-342-647-27	Sequence 27, Appl	c 361	33	61.1	7676	4	US-09-056-556-213	Sequence 213, App
c 289	33	61.1	2387	4	US-09-634-957-1	Sequence 1, Appli	c 362	33	61.1	7676	4	US-09-072-596-208	Sequence 208, App
c 290	33	61.1	2387	4	US-09-723-153-1	Sequence 1, Appli	c 363	33	61.1	7705	2	US-08-687-080-115	Sequence 115, App
c 291	33	61.1	2387	4	US-09-723-429-1	Sequence 1, Appli	c 364	33	61.1	9306	4	US-09-453-702B-231	Sequence 231, App
c 292	33	61.1	2412	4	US-09-072-529-345	Sequence 345, App	c 365	33	61.1	9600	4	US-08-910-647-1	Sequence 1, Appli
c 293	33	61.1	2573	2	US-08-884-681-2	Sequence 2, Appli	c 366	33	61.1	9600	4	US-09-620-925-1	Sequence 1, Appli
c 294	33	61.1	2573	4	US-09-258-643-2	Sequence 2, Appli	c 367	33	61.1	10240	4	US-08-961-527-32	Sequence 32, Appl
c 295	33	61.1	2580	3	US-09-050-863-2	Sequence 2, Appli	c 368	33	61.1	10342	4	US-08-972-927-5	Sequence 5, Appli
c 296	33	61.1	2580	4	US-09-359-081-2	Sequence 2, Appli	c 369	33	61.1	10596	1	US-07-884-811-15	Sequence 15, Appl
c 297	33	61.1	2739	2	US-08-658-665-59	Sequence 59, Appl	c 370	33	61.1	10596	1	US-07-885-971-15	Sequence 15, Appl
c 298	33	61.1	2739	4	US-08-796-101-23	Sequence 23, Appl	c 371	33	61.1	10596	1	US-08-087-783A-15	Sequence 15, Appl
c 299	33	61.1	2739	4	US-09-085-273-59	Sequence 59, Appl	c 372	33	61.1	10596	1	US-08-194-088B-15	Sequence 15, Appl
c 300	33	61.1	3014	1	US-08-629-939-1	Sequence 1, Appli	c 373	33	61.1	10596	2	US-08-194-087-15	Sequence 15, Appl
c 301	33	61.1	3014	1	US-08-759-873-1	Sequence 1, Appli	c 374	33	61.1	10596	5	PCT-US93-04648-15	Sequence 15, Appl
c 302	33	61.1	3245	1	US-07-935-311A-3	Sequence 3, Appli	c 375	33	61.1	15297	4	US-09-817-180-3	Sequence 3, Appli
c 303	33	61.1	3245	1	US-08-368-079-3	Sequence 3, Appli	c 376	33	61.1	15936	4	US-09-147-119-1	Sequence 1, Appli
c 304	33	61.1	3245	5	PCT-US93-07996-3	Sequence 3, Appli	c 377	33	61.1	17949	4	US-09-087-465-3	Sequence 3, Appli

378	33	61.1	24183	4	US-08-943-731-3	Sequence 3, Appl	451	32	59.3	553	3	US-08-370-223-3	Sequence 3, Appl
379	33	61.1	30001	1	US-08-125-468-1	Sequence 1, Appl	452	32	59.3	571	4	US-09-404-879A-48	Sequence 48, Appl
380	33	61.1	30001	2	US-08-474-933-1	Sequence 1, Appl	453	32	59.3	573	1	US-08-638-911A-4	Sequence 4, Appl
381	33	61.1	34446	4	US-09-103-330-35	Sequence 35, Appl	454	32	59.3	573	1	US-08-638-911A-54	Sequence 54, Appl
382	33	61.1	37950	4	US-09-338-907-183	Sequence 183, Appl	C 455	32	59.3	578	1	US-09-812-484-13	Sequence 13, Appl
383	33	61.1	37950	4	US-09-218-207-183	Sequence 183, Appl	C 456	32	59.3	583	1	US-08-652-127C-8	Sequence 8, Appl
384	33	61.1	46899	3	US-08-471-118A-1	Sequence 1, Appl	C 457	32	59.3	586	1	US-08-466-033-42	Sequence 42, Appl
385	33	61.1	68750	1	US-09-335-409-1	Sequence 1, Appl	458	32	59.3	596	2	US-08-444-733-42	Sequence 42, Appl
386	33	61.1	68750	4	US-09-568-102-1	Sequence 1, Appl	459	32	59.3	596	2	US-08-464-134-42	Sequence 42, Appl
387	33	61.1	68750	4	US-09-567-969-1	Sequence 1, Appl	460	32	59.3	596	2	US-08-461-361-42	Sequence 42, Appl
388	33	61.1	68750	4	US-09-568-480-1	Sequence 1, Appl	461	32	59.3	596	2	US-08-485-910-42	Sequence 42, Appl
389	33	61.1	68750	4	US-09-568-486-1	Sequence 1, Appl	462	32	59.3	610	3	US-08-906-769-140	Sequence 140, App
390	33	61.1	68750	4	US-09-568-472-1	Sequence 1, Appl	463	32	59.3	610	3	US-08-906-616-140	Sequence 140, App
391	33	61.1	68750	4	US-09-567-899-1	Sequence 1, Appl	464	32	59.3	610	3	US-08-639-075A-140	Sequence 140, App
392	33	61.1	71989	4	US-09-443-501A-2	Sequence 2, Appl	465	32	59.3	610	4	US-09-012-431-140	Sequence 140, App
393	33	61.1	81001	4	US-09-750-580-1	Sequence 1, Appl	466	32	59.3	610	4	US-09-012-692-140	Sequence 140, App
394	33	61.1	81001	4	US-09-750-580-1	Sequence 1, Appl	467	32	59.3	610	4	US-08-906-613-140	Sequence 140, App
395	33	61.1	246240	2	US-08-724-394A-20	Sequence 20, Appl	468	32	59.3	622	3	US-08-545-809A-59	Sequence 59, Appl
396	33	61.1	246240	2	US-08-724-394A-20	Sequence 20, Appl	469	32	59.3	633	2	US-08-969-106-12	Sequence 12, Appl
397	33	61.1	246240	2	US-08-724-394A-21	Sequence 21, Appl	470	32	59.3	650	3	US-08-545-809A-4	Sequence 4, Appl
398	33	61.1	246240	2	US-08-724-394A-21	Sequence 21, Appl	471	32	59.3	661	4	US-09-328-111-69	Sequence 69, Appl
399	33	61.1	246240	2	US-08-724-394A-22	Sequence 22, Appl	472	32	59.3	694	4	US-09-605-785-734	Sequence 734, App
400	33	61.1	246240	2	US-08-724-394A-22	Sequence 22, Appl	C 473	32	59.3	704	2	US-08-602-093-5	Sequence 5, Appl
401	33	61.1	4411529	4	US-09-103-840A-1	Sequence 1, Appl	474	32	59.3	752	1	US-08-725-182C-1	Sequence 1, Appl
402	32.5	60.2	2574	4	US-09-142-529-2	Sequence 2, Appl	475	32	59.3	752	3	US-09-013-150-1	Sequence 1, Appl
403	32	59.3	58	2	US-08-467-948A-20	Sequence 20, Appl	476	32	59.3	752	4	US-09-359-503-1	Sequence 1, Appl
404	32	59.3	58	2	US-08-467-947A-20	Sequence 20, Appl	477	32	59.3	752	4	US-09-062-422-1	Sequence 1, Appl
405	32	59.3	71	1	US-07-805-567-16	Sequence 16, Appl	478	32	59.3	752	4	US-09-396-184-1	Sequence 1, Appl
406	32	59.3	71	1	US-08-105-483-109	Sequence 109, App	479	32	59.3	752	4	US-08-937-263B-1	Sequence 1, Appl
407	32	59.3	71	1	US-08-220-151-66	Sequence 66, Appl	480	32	59.3	755	1	US-08-791-495-6	Sequence 6, Appl
408	32	59.3	71	1	US-08-413-118-66	Sequence 66, Appl	481	32	59.3	755	1	US-08-791-495-8	Sequence 8, Appl
409	32	59.3	71	1	US-08-224-391-62	Sequence 62, Appl	C 482	32	59.3	796	4	US-08-927-219-36	Sequence 36, Appl
410	32	59.3	71	1	US-08-484-304-62	Sequence 62, Appl	C 483	32	59.3	836	4	US-09-605-785-733	Sequence 36, Appl
411	32	59.3	71	1	US-08-224-657-43	Sequence 43, Appl	C 484	32	59.3	836	4	US-08-969-106-10	Sequence 10, Appl
412	32	59.3	71	1	US-08-709-209-109	Sequence 109, App	C 485	32	59.3	944	6	5223391-6	Patent No. 5223391
413	32	59.3	71	1	US-08-458-101-109	Sequence 101, App	C 486	32	59.3	981	2	US-08-841-349-15	Sequence 15, Appl
414	32	59.3	71	1	US-08-184-009-42	Sequence 42, Appl	C 487	32	59.3	981	2	US-09-134-001C-982	Sequence 982, App
415	32	59.3	71	2	US-08-486-369-31	Sequence 31, Appl	C 488	32	59.3	1002	1	US-08-791-495-4	Sequence 4, Appl
416	32	59.3	71	2	US-08-417-210A-42	Sequence 42, Appl	C 489	32	59.3	1002	4	US-09-625-040-4	Sequence 4, Appl
417	32	59.3	71	2	US-08-458-356-42	Sequence 42, Appl	C 490	32	59.3	1032	1	US-08-599-171A-20	Sequence 20, Appl
418	32	59.3	71	2	US-08-471-025-31	Sequence 31, Appl	C 491	32	59.3	1032	2	US-08-646-590B-20	Sequence 20, Appl
419	32	59.3	71	3	US-08-658-665-31	Sequence 31, Appl	C 492	32	59.3	1032	2	US-09-069-226-20	Sequence 20, Appl
420	32	59.3	71	3	US-08-473-446-66	Sequence 66, Appl	C 493	32	59.3	1032	4	US-09-412-184-20	Sequence 20, Appl
421	32	59.3	71	4	US-08-460-736-42	Sequence 42, Appl	C 494	32	59.3	1059	4	US-09-576-160B-7	Sequence 7, Appl
422	32	59.3	71	4	US-09-085-273-31	Sequence 31, Appl	C 495	32	59.3	1071	2	US-08-997-080-180	Sequence 180, App
423	32	59.3	71	4	US-09-354-138-43	Sequence 43, Appl	C 496	32	59.3	1071	2	US-08-997-362-180	Sequence 180, App
424	32	59.3	71	5	PCT-US96-00547-31	Sequence 31, Appl	C 497	32	59.3	1071	4	US-09-322-478-25	Sequence 25, Appl
425	32	59.3	92	4	US-08-943-731-66	Sequence 66, Appl	C 498	32	59.3	1071	4	US-09-322-478-25	Sequence 25, Appl
426	32	59.3	193	4	US-08-817-787-16	Sequence 16, Appl	C 499	32	59.3	1071	4	US-09-322-478-25	Sequence 25, Appl
427	32	59.3	201	2	US-08-592-383-6	Sequence 6, Appl	C 500	32	59.3	1080	4	US-09-134-001C-1654	Sequence 1654, Ap
428	32	59.3	235	2	US-08-997-080-138	Sequence 138, App	501	32	59.3	1116	4	US-09-104-308-2	Sequence 2, Appl
429	32	59.3	235	2	US-08-997-362-138	Sequence 138, App	502	32	59.3	1116	4	US-09-321-981-2	Sequence 21, Appl
430	32	59.3	235	2	US-09-095-855-138	Sequence 138, App	503	32	59.3	1132	3	US-08-651-136C-21	Sequence 21, Appl
431	32	59.3	235	4	US-09-324-542-138	Sequence 138, App	504	32	59.3	1132	4	US-09-229-911A-21	Sequence 21, Appl
432	32	59.3	235	4	US-09-205-426-138	Sequence 138, App	505	32	59.3	1212	3	US-08-545-809A-61	Sequence 61, Appl
433	32	59.3	253	1	US-08-507-615-1	Sequence 1, Appl	506	32	59.3	1254	4	US-09-425-578-3	Sequence 3, Appl
434	32	59.3	261	4	US-09-060-756-102	Sequence 102, App	507	32	59.3	1309	4	US-09-433-248A-1	Sequence 1, Appl
435	32	59.3	288	4	US-08-851-362D-15	Sequence 15, Appl	C 508	32	59.3	1386	4	US-09-287-097-1	Sequence 1, Appl
436	32	59.3	291	4	US-08-851-362D-13	Sequence 13, Appl	C 509	32	59.3	1406	4	US-09-625-040-1	Sequence 1, Appl
437	32	59.3	326	1	US-08-466-033-105	Sequence 105, App	510	32	59.3	1448	4	US-08-235-477B-1	Sequence 1, Appl
438	32	59.3	326	1	US-08-444-733-105	Sequence 105, App	511	32	59.3	1520	5	PCT-US95-04353-1	Sequence 1, Appl
439	32	59.3	326	2	US-08-464-134-105	Sequence 105, App	C 512	32	59.3	1520	5	US-08-881-857-1	Sequence 1, Appl
440	32	59.3	326	2	US-08-461-361-105	Sequence 105, App	C 513	32	59.3	1556	2	US-09-233-342A-1	Sequence 1, Appl
441	32	59.3	326	5	PCT-US95-06266-88	Sequence 88, Appl	C 514	32	59.3	1556	2	US-08-405-175A-10	Sequence 10, Appl
442	32	59.3	342	2	US-08-245-511-17	Sequence 17, Appl	C 515	32	59.3	1566	4	US-08-552-369-2	Sequence 2, Appl
443	32	59.3	342	2	US-08-600-993A-17	Sequence 17, Appl	C 516	32	59.3	1575	2	US-09-673-018-1	Sequence 1, Appl
444	32	59.3	354	2	US-08-652-816A-23	Sequence 23, Appl	C 517	32	59.3	1586	4	US-08-492-459-17	Sequence 17, Appl
445	32	59.3	372	2	US-08-477-553A-46	Sequence 46, Appl	C 518	32	59.3	1601	3	US-08-492-459-18	Sequence 18, Appl
446	32	59.3	407	4	US-09-056-556-173	Sequence 173, App	C 519	32	59.3	1601	3	US-08-423-752-17	Sequence 17, Appl
447	32	59.3	407	4	US-09-072-596-168	Sequence 168, App	C 520	32	59.3	1601	3	US-08-423-752-18	Sequence 18, Appl
448	32	59.3	426	2	US-08-480-774A-1	Sequence 1, Appl	C 521	32	59.3	1601	3	US-08-423-752-18	Sequence 18, Appl
449	32	59.3	426	2	US-08-480-774A-1	Sequence 1, Appl	C 522	32	59.3	1601	3	US-08-423-752-18	Sequence 18, Appl
450	32	59.3	486	4	US-09-149-476-32	Sequence 32, Appl	C 523	32	59.3	1601	4	US-08-716-873-31	Sequence 31, Appl

c 524	32	59.3	1601	4	US-08-716-873-32	Sequence 32, Appl	c 597	32	59.3	2928	5	PCT-US92-03320A-3	Sequence 3, Appl
c 525	32	59.3	1601	4	US-09-368-431-31	Sequence 31, Appl	c 598	32	59.3	2940	6	US-08-592-383-1	Sequence 1, Appl
c 526	32	59.3	1601	4	US-09-368-431-32	Sequence 32, Appl	c 599	32	59.3	2940	6	5171671-1	Patent No. 5171671
c 527	32	59.3	1601	4	US-09-414-006-17	Sequence 17, Appl	600	32	59.3	2961	2	US-08-407-875-1	Sequence 1, Appl
c 528	32	59.3	1601	4	US-09-414-006-18	Sequence 18, Appl	601	32	59.3	2961	4	US-09-277-858-1	Sequence 1, Appl
c 529	32	59.3	1601	4	US-09-447-223-17	Sequence 17, Appl	c 602	32	59.3	3036	1	US-08-306-691B-52	Sequence 52, Appl
c 530	32	59.3	1601	4	US-09-447-223-18	Sequence 18, Appl	c 603	32	59.3	3036	2	US-08-095-728B-1	Sequence 1, Appl
c 531	32	59.3	1710	2	US-08-480-736-1	Sequence 1, Appl	c 604	32	59.3	3036	5	PCT-US92-03320A-1	Sequence 1, Appl
c 532	32	59.3	1716	4	US-09-321-981-4	Sequence 4, Appl	c 605	32	59.3	3095	4	US-09-325-932A-128	Sequence 128, App
c 533	32	59.3	1766	4	US-09-182-145-9	Sequence 9, Appl	606	32	59.3	3164	1	US-07-723-002C-3	Sequence 3, Appl
c 534	32	59.3	1766	4	US-09-182-145-10	Sequence 10, Appl	c 607	32	59.3	3230	4	US-08-961-527-203	Sequence 203, App
c 535	32	59.3	1829	1	US-07-688-352C-15	Sequence 15, Appl	608	32	59.3	3260	1	US-08-049-254-1	Sequence 1, Appl
c 536	32	59.3	1829	2	US-08-474-379C-15	Sequence 15, Appl	609	32	59.3	3260	1	US-08-472-934-1	Sequence 1, Appl
c 537	32	59.3	1829	3	US-09-146-249A-15	Sequence 15, Appl	610	32	59.3	3260	2	US-08-323-460A-1	Sequence 1, Appl
c 538	32	59.3	1829	3	US-08-206-188B-15	Sequence 15, Appl	611	32	59.3	3260	2	US-08-461-146C-1	Sequence 1, Appl
c 539	32	59.3	1829	5	PCT-US91-02714-15	Sequence 15, Appl	612	32	59.3	3260	3	US-08-461-145C-1	Sequence 1, Appl
c 540	32	59.3	1848	1	US-08-638-911A-34	Sequence 34, Appl	613	32	59.3	3260	3	US-08-628-829-1	Sequence 1, Appl
c 541	32	59.3	1849	2	US-08-676-166A-1	Sequence 1, Appl	614	32	59.3	3340	4	US-09-021-560-1	Sequence 1, Appl
c 542	32	59.3	1878	1	US-07-612-673-1	Sequence 1, Appl	c 615	32	59.3	3396	2	US-08-974-565C-6	Sequence 6, Appl
c 543	32	59.3	1879	1	US-07-796-361A-10	Sequence 10, Appl	c 616	32	59.3	3396	3	US-09-255-748-6	Sequence 6, Appl
c 544	32	59.3	1879	1	US-08-539-666-1	Sequence 1, Appl	617	32	59.3	3480	4	US-09-226-012-1	Sequence 1, Appl
c 545	32	59.3	1908	1	US-08-173-508-1	Sequence 1, Appl	c 618	32	59.3	3511	3	US-08-892-747-13	Sequence 13, Appl
c 546	32	59.3	1908	2	US-08-265-310-1	Sequence 1, Appl	c 619	32	59.3	3554	2	US-08-460-309-1	Sequence 1, Appl
c 547	32	59.3	1908	3	US-08-951-742-1	Sequence 1, Appl	c 620	32	59.3	3554	2	US-08-125-077-1	Sequence 1, Appl
c 548	32	59.3	1967	1	US-08-638-911A-51	Sequence 51, Appl	c 621	32	59.3	3554	6	544158-1	Patent No. 544158
c 549	32	59.3	1988	2	US-08-712-948-8	Sequence 8, Appl	c 622	32	59.3	3708	2	US-08-680-326-29	Sequence 29, Appl
c 550	32	59.3	1996	3	US-08-946-026-15	Sequence 15, Appl	623	32	59.3	3911	4	US-09-423-890-1	Sequence 1, Appl
c 551	32	59.3	2051	1	US-09-199-637A-114	Sequence 114, App	624	32	59.3	3950	4	US-09-226-012-3	Sequence 3, Appl
c 552	32	59.3	2070	1	US-08-233-146-7	Sequence 7, Appl	c 625	32	59.3	3973	2	US-08-602-093-6	Sequence 6, Appl
c 553	32	59.3	2070	1	US-08-028-463-13	Sequence 13, Appl	c 626	32	59.3	4088	2	US-08-317-310A-1	Sequence 1, Appl
c 554	32	59.3	2070	1	US-08-461-836-13	Sequence 13, Appl	c 627	32	59.3	4088	5	PCT-US95-13041-1	Sequence 1, Appl
c 555	32	59.3	2070	1	US-08-463-470-7	Sequence 7, Appl	628	32	59.3	4105	4	US-08-121-446-1	Sequence 1, Appl
c 556	32	59.3	2196	3	US-09-142-759-2	Sequence 2, Appl	c 629	32	59.3	4615	2	US-08-674-351-3	Sequence 3, Appl
c 557	32	59.3	2196	4	US-09-360-545-1	Sequence 1, Appl	630	32	59.3	5091	4	US-08-469-260A-668	Sequence 668, App
c 558	32	59.3	2196	4	US-09-398-395A-29	Sequence 29, Appl	c 631	32	59.3	5199	3	US-08-726-214-13	Sequence 13, Appl
c 559	32	59.3	2201	2	US-08-974-565C-2	Sequence 2, Appl	c 632	32	59.3	5253	4	US-09-423-890-7	Sequence 7, Appl
c 560	32	59.3	2201	3	US-09-255-748-2	Sequence 2, Appl	c 633	32	59.3	5288	2	US-08-540-406-18	Sequence 18, Appl
c 561	32	59.3	2205	4	US-09-360-545-31	Sequence 31, Appl	c 634	32	59.3	5288	3	US-08-656-055-18	Sequence 18, Appl
c 562	32	59.3	2224	4	US-09-261-855-1	Sequence 1, Appl	c 635	32	59.3	5288	4	US-08-954-668-18	Sequence 18, Appl
c 563	32	59.3	2229	1	US-08-624-663A-1	Sequence 1, Appl	c 636	32	59.3	5288	4	US-08-918-658-18	Sequence 18, Appl
c 564	32	59.3	2232	1	US-08-638-911A-53	Sequence 53, Appl	c 637	32	59.3	5288	5	PCT-US95-13233-18	Sequence 18, Appl
c 565	32	59.3	2309	4	US-08-556-627A-1	Sequence 1, Appl	c 638	32	59.3	5539	4	US-08-628-829-3	Sequence 3, Appl
c 566	32	59.3	2320	4	US-09-813-817-1	Sequence 1, Appl	639	32	59.3	5658	4	US-08-881-450A-23	Sequence 23, Appl
c 567	32	59.3	2320	4	US-09-978-197-1	Sequence 1, Appl	640	32	59.3	6089	4	US-09-221-017B-8	Sequence 8, Appl
c 568	32	59.3	2344	3	US-08-893-852A-2	Sequence 2, Appl	c 641	32	59.3	6170	2	US-08-494-907-4	Sequence 4, Appl
c 569	32	59.3	2377	4	US-08-556-627A-3	Sequence 3, Appl	c 642	32	59.3	6170	5	PCT-US96-10986-4	Sequence 4, Appl
c 570	32	59.3	2379	1	US-08-638-911A-52	Sequence 52, Appl	c 643	32	59.3	6387	2	US-08-494-907-3	Sequence 3, Appl
c 571	32	59.3	2385	3	US-08-432-459-1	Sequence 1, Appl	c 644	32	59.3	6387	5	PCT-US96-10986-3	Sequence 3, Appl
c 572	32	59.3	2385	3	US-08-432-459-3	Sequence 3, Appl	c 645	32	59.3	6563	4	US-09-453-702B-192	Sequence 192, App
c 573	32	59.3	2385	3	US-08-423-752-1	Sequence 1, Appl	646	32	59.3	6749	4	US-08-961-527-84	Sequence 84, Appl
c 574	32	59.3	2385	3	US-08-423-752-3	Sequence 3, Appl	c 647	32	59.3	7198	4	US-08-994-035C-4	Sequence 4, Appl
c 575	32	59.3	2385	4	US-08-716-873-15	Sequence 15, Appl	c 648	32	59.3	7198	4	US-09-395-861-4	Sequence 4, Appl
c 576	32	59.3	2385	4	US-08-716-873-17	Sequence 17, Appl	c 649	32	59.3	7542	4	US-09-734-030-3	Sequence 3, Appl
c 577	32	59.3	2385	4	US-09-368-431-15	Sequence 15, Appl	650	32	59.3	7868	6	5175383-1	Patent No. 5175383
c 578	32	59.3	2385	4	US-09-368-431-17	Sequence 17, Appl	c 651	32	59.3	7868	2	US-08-751-189-2	Sequence 2, Appl
c 579	32	59.3	2385	4	US-09-414-006-1	Sequence 1, Appl	c 652	32	59.3	7886	2	US-09-060-836-2	Sequence 2, Appl
c 580	32	59.3	2385	4	US-09-414-006-3	Sequence 3, Appl	c 653	32	59.3	7886	4	US-09-184-445-2	Sequence 2, Appl
c 581	32	59.3	2385	4	US-09-447-223-1	Sequence 1, Appl	654	32	59.3	8236	1	US-08-461-837-1	Sequence 1, Appl
c 582	32	59.3	2389	2	US-09-447-223-3	Sequence 3, Appl	655	32	59.3	8236	3	US-08-973-223-1	Sequence 1, Appl
c 583	32	59.3	2399	4	US-08-969-106-9	Sequence 9, Appl	656	32	59.3	8236	4	US-09-347-060-1	Sequence 1, Appl
c 584	32	59.3	2461	1	US-08-282-141-1	Sequence 1, Appl	657	32	59.3	8236	5	PCT-US96-09495-1	Sequence 1, Appl
c 585	32	59.3	2528	1	US-08-439-131A-1	Sequence 1, Appl	658	32	59.3	8298	5	PCT-US93-03076-1	Sequence 1, Appl
c 586	32	59.3	2528	1	US-08-440-674-1	Sequence 1, Appl	c 659	32	59.3	8655	3	US-09-075-272-1	Sequence 1, Appl
c 587	32	59.3	2528	1	US-08-107-348-1	Sequence 1, Appl	660	32	59.3	8878	1	US-08-759-444-2	Sequence 2, Appl
c 588	32	59.3	2561	1	US-08-638-911A-24	Sequence 24, Appl	661	32	59.3	9034	4	US-08-469-260A-397	Sequence 397, App
c 589	32	59.3	2561	1	US-08-638-911A-26	Sequence 26, Appl	662	32	59.3	9103	1	US-08-466-033-182	Sequence 182, App
c 590	32	59.3	2635	3	US-09-126-280-3	Sequence 3, Appl	663	32	59.3	9103	2	US-08-444-733-182	Sequence 182, App
c 591	32	59.3	2635	3	US-09-126-280-3	Sequence 3, Appl	664	32	59.3	9103	2	US-08-464-134-182	Sequence 182, App
c 592	32	59.3	2658	2	US-08-592-383-3	Sequence 3, Appl	665	32	59.3	9103	2	US-08-461-361-182	Sequence 182, App
c 593	32	59.3	2670	3	US-09-126-280-3	Sequence 3, Appl	666	32	59.3	9103	2	US-08-485-910-182	Sequence 182, App
c 594	32	59.3	2717	1	US-07-723-002C-1	Sequence 1, Appl	667	32	59.3	9103	5	PCT-US95-06266-156	Sequence 156, App
c 595	32	59.3	2759	4	US-09-144-367-1	Sequence 1, Appl	668	32	59.3	9122	2	US-08-417-629B-1	Sequence 1, Appl
c 596	32	59.3	2928	2	US-08-095-728B-3	Sequence 3, Appl	669	32	59.3	9126	1	US-08-580-038-26	Sequence 26, Appl

670	32	59.3	9126	2	US-08-639-857-3	Sequence 3, Appl	c 743	31	57.4	285	1	US-08-777-019-25	Sequence 25, Appl
671	32	59.3	9327	1	US-08-466-033-234	Sequence 234, App	c 744	31	57.4	285	2	US-08-777-143-25	Sequence 25, Appl
672	32	59.3	9327	1	US-08-444-733-234	Sequence 234, App	c 745	31	57.4	285	3	US-08-775-414-25	Sequence 25, Appl
673	32	59.3	9327	2	US-08-464-134-234	Sequence 234, App	c 746	31	57.4	285	4	US-08-931-858E-25	Sequence 25, Appl
674	32	59.3	9327	2	US-08-461-361-234	Sequence 234, App	c 747	31	57.4	285	4	US-08-981-739-25	Sequence 25, Appl
675	32	59.3	9327	2	US-08-485-910-234	Sequence 234, App	c 748	31	57.4	285	4	US-09-128-026-25	Sequence 25, Appl
676	32	59.3	9391	5	US-08-638-911A-1	Sequence 1, Appl	749	31	57.4	307	4	US-09-397-787-174	Sequence 174, App
677	32	59.3	9391	5	PCT-US95-06266-14	Sequence 14, Appl	750	31	57.4	311	2	US-09-060-288-3	Sequence 3, Appl
678	32	59.3	9392	1	US-08-466-033-14	Sequence 14, Appl	751	31	57.4	311	4	US-08-872-571A-3	Sequence 3, Appl
679	32	59.3	9392	2	US-08-444-733-14	Sequence 14, Appl	752	31	57.4	311	4	US-08-274-608-10	Sequence 10, Appl
680	32	59.3	9392	2	US-08-464-134-14	Sequence 14, Appl	753	31	57.4	311	4	US-08-377-891-3	Sequence 3, Appl
681	32	59.3	9392	2	US-08-461-361-14	Sequence 14, Appl	754	31	57.4	311	4	US-08-771-190-10	Sequence 10, Appl
682	32	59.3	9392	2	US-08-485-910-14	Sequence 14, Appl	755	31	57.4	311	5	PCT-US94-07955-10	Sequence 10, Appl
683	32	59.3	9775	4	US-08-977-171-1	Sequence 1, Appl	c 756	31	57.4	317	4	US-08-991-789A-181	Sequence 181, App
684	32	59.3	9880	3	US-08-680-897-1	Sequence 1, Appl	c 757	31	57.4	317	4	US-09-062-451-181	Sequence 181, App
685	32	59.3	9934	4	US-08-977-171-2	Sequence 2, Appl	c 758	31	57.4	317	4	US-09-598-326-181	Sequence 181, App
686	32	59.3	10014	4	US-08-927-219-130	Sequence 130, App	c 759	31	57.4	344	3	US-08-370-223-10	Sequence 10, Appl
687	32	59.3	10684	3	US-08-618-1008-3	Sequence 3, Appl	c 760	31	57.4	348	1	US-08-519-777-13	Sequence 13, Appl
688	32	59.3	11517	1	US-07-920-281C-1	Sequence 1, Appl	c 761	31	57.4	348	1	US-08-742-035-13	Sequence 13, Appl
689	32	59.3	11517	4	US-08-466-277-1	Sequence 1, Appl	c 762	31	57.4	348	1	US-08-777-019-13	Sequence 13, Appl
690	32	59.3	13206	4	US-08-961-527-33	Sequence 33, Appl	c 763	31	57.4	348	2	US-08-777-143-13	Sequence 13, Appl
691	32	59.3	14636	4	US-09-173-914-6	Sequence 6, Appl	c 764	31	57.4	348	3	US-08-775-414-13	Sequence 13, Appl
692	32	59.3	15079	4	US-09-385-028-1	Sequence 1, Appl	c 765	31	57.4	348	4	US-08-931-858E-13	Sequence 13, Appl
693	32	59.3	15101	2	US-08-799-464A-14	Sequence 14, Appl	c 766	31	57.4	348	4	US-08-981-739-13	Sequence 13, Appl
694	32	59.3	15101	5	PCT-US95-09927-14	Sequence 14, Appl	c 767	31	57.4	357	4	US-09-128-026-13	Sequence 13, Appl
695	32	59.3	15108	4	US-08-157-005-1	Sequence 1, Appl	c 768	31	57.4	357	4	US-09-417-455-1	Sequence 1, Appl
696	32	59.3	15108	4	US-08-747-863-1	Sequence 1, Appl	c 769	31	57.4	357	4	US-09-280-116-131	Sequence 131, App
697	32	59.3	15108	4	US-09-565-864-1	Sequence 1, Appl	c 770	31	57.4	357	4	US-09-348-942-1	Sequence 1, Appl
698	32	59.3	15328	2	US-08-988-497-33	Sequence 33, Appl	c 771	31	57.4	357	4	US-09-457-626-1	Sequence 1, Appl
699	32	59.3	15328	4	US-09-362-230-33	Sequence 33, Appl	c 772	31	57.4	361	6	5304466-1	Patent No. 5304466
700	32	59.3	15328	5	PCT-US94-70926-33	Sequence 33, Appl	c 773	31	57.4	400	4	US-09-056-556-179	Sequence 179, App
701	32	59.3	25165	4	US-09-453-702B-39	Sequence 39, Appl	c 774	31	57.4	400	4	US-09-072-596-174	Sequence 174, App
702	32	59.3	35828	4	US-09-449-2180-17	Sequence 17, Appl	c 775	31	57.4	409	4	US-09-174-768-1	Sequence 1, Appl
703	32	59.3	38564	4	US-09-734-673-3	Sequence 3, Appl	c 776	31	57.4	417	4	US-09-283-144-1	Sequence 1, Appl
704	32	59.3	38682	4	US-08-943-731-2	Sequence 2, Appl	c 777	31	57.4	440	4	US-09-397-787-328	Sequence 328, App
705	32	59.3	61663	4	US-09-453-702B-62	Sequence 62, Appl	c 778	31	57.4	458	4	US-09-605-785-745	Sequence 745, App
706	32	59.3	162450	4	US-09-345-882-1	Sequence 1, Appl	c 779	31	57.4	484	4	US-09-370-838-283	Sequence 283, App
707	31.5	58.3	93	3	US-08-976-413A-403	Sequence 403, App	c 780	31	57.4	489	4	US-09-527-236A-10	Sequence 10, Appl
708	31	57.4	31	2	US-08-589-756-5	Sequence 5, Appl	c 781	31	57.4	490	1	US-08-488-961-1	Sequence 1, Appl
709	31	57.4	31	4	US-09-206-800-5	Sequence 5, Appl	c 782	31	57.4	490	4	US-08-973-297-1	Sequence 1, Appl
710	31	57.4	31	4	US-09-206-898-5	Sequence 5, Appl	c 783	31	57.4	490	5	PCT-US96-06511-1	Sequence 1, Appl
711	31	57.4	40	4	US-09-189-462-25	Sequence 25, Appl	c 784	31	57.4	491	4	US-09-643-597-307	Sequence 307, App
712	31	57.4	40	4	US-09-189-462-26	Sequence 26, Appl	c 785	31	57.4	507	1	US-08-488-961-11	Sequence 11, Appl
713	31	57.4	41	4	US-09-222-939-51	Sequence 51, Appl	c 786	31	57.4	507	1	US-08-973-297-11	Sequence 11, Appl
714	31	57.4	60	1	US-08-633-779-2	Sequence 2, Appl	c 787	31	57.4	507	5	PCT-US96-06511-11	Sequence 11, Appl
715	31	57.4	70	4	US-09-364-380-27	Sequence 27, Appl	c 788	31	57.4	519	3	US-09-026-343-22	Sequence 22, Appl
716	31	57.4	100	3	US-08-441-971-24	Sequence 24, Appl	c 789	31	57.4	519	4	US-09-362-871-22	Sequence 22, Appl
717	31	57.4	100	4	US-08-221-653-24	Sequence 24, Appl	c 790	31	57.4	542	3	US-09-058-489-87	Sequence 87, Appl
718	31	57.4	100	4	US-08-442-144A-24	Sequence 24, Appl	c 791	31	57.4	552	3	US-09-058-489-41	Sequence 41, Appl
719	31	57.4	100	4	US-08-441-970-24	Sequence 24, Appl	c 792	31	57.4	591	1	US-08-519-777-11	Sequence 11, Appl
720	31	57.4	154	4	US-09-280-116-145	Sequence 145, App	c 793	31	57.4	591	1	US-08-742-035-11	Sequence 11, Appl
721	31	57.4	169	1	US-08-519-777-27	Sequence 27, Appl	c 794	31	57.4	591	2	US-08-777-019-11	Sequence 11, Appl
722	31	57.4	169	1	US-08-742-035-27	Sequence 27, Appl	c 795	31	57.4	591	2	US-08-777-143-11	Sequence 11, Appl
723	31	57.4	169	1	US-08-777-019-27	Sequence 27, Appl	c 796	31	57.4	591	3	US-08-775-414-11	Sequence 11, Appl
724	31	57.4	169	2	US-08-777-143-27	Sequence 27, Appl	c 797	31	57.4	591	4	US-08-931-858E-11	Sequence 11, Appl
725	31	57.4	169	3	US-08-775-414-27	Sequence 27, Appl	c 798	31	57.4	591	4	US-08-981-739-11	Sequence 11, Appl
726	31	57.4	169	4	US-08-931-858E-27	Sequence 27, Appl	c 799	31	57.4	595	4	US-09-128-026-11	Sequence 11, Appl
727	31	57.4	169	4	US-08-981-739-27	Sequence 27, Appl	c 800	31	57.4	595	4	US-09-385-982-282	Sequence 282, App
728	31	57.4	169	4	US-09-128-026-27	Sequence 27, Appl	c 801	31	57.4	615	4	US-09-328-111-371	Sequence 371, App
729	31	57.4	227	4	US-09-056-556-165	Sequence 165, App	c 802	31	57.4	634	4	US-09-470-191-84	Sequence 84, Appl
730	31	57.4	227	4	US-09-072-596-160	Sequence 160, App	c 803	31	57.4	690	4	US-09-404-879A-321	Sequence 321, App
731	31	57.4	228	1	US-08-519-777-20	Sequence 20, App	c 804	31	57.4	695	2	US-08-403-852D-7	Sequence 7, Appl
732	31	57.4	228	1	US-08-742-035-20	Sequence 20, App	c 805	31	57.4	695	3	US-08-510-646B-7	Sequence 7, Appl
733	31	57.4	228	1	US-08-777-019-20	Sequence 20, App	c 806	31	57.4	695	4	US-09-231-818-7	Sequence 7, Appl
734	31	57.4	228	2	US-08-777-143-20	Sequence 20, App	c 807	31	57.4	703	3	US-09-058-489-88	Sequence 88, Appl
735	31	57.4	228	3	US-08-775-414-20	Sequence 20, App	c 808	31	57.4	704	2	US-08-998-416-198	Sequence 198, App
736	31	57.4	228	4	US-08-931-858E-20	Sequence 20, App	c 809	31	57.4	704	2	US-08-602-093-5	Sequence 5, Appl
737	31	57.4	228	4	US-08-981-739-20	Sequence 20, App	c 810	31	57.4	711	3	US-08-622-046B-2	Sequence 2, Appl
738	31	57.4	228	4	US-09-128-026-20	Sequence 20, App	c 811	31	57.4	711	3	US-08-622-046B-13	Sequence 13, Appl
739	31	57.4	239	5	PCT-US93-08106-7	Sequence 7, Appl	c 812	31	57.4	713	4	US-09-100-264-2	Sequence 2, Appl
740	31	57.4	239	5	PCT-US94-00089-15	Sequence 15, Appl	c 813	31	57.4	713	4	US-08-998-416-1098	Sequence 1098, Ap
741	31	57.4	285	1	US-08-519-777-25	Sequence 25, Appl	c 814	31	57.4	721	1	US-08-229-515A-15	Sequence 15, Appl
742	31	57.4	285	1	US-08-742-035-25	Sequence 25, Appl	c 815	31	57.4	721	1	US-08-645-865-15	Sequence 15, Appl



816	31	57.4	750	3	US-08-768-859A-7	Sequence 7, Appli	c 889	31	57.4	1000	1	US-07-960-112B-1	Sequence 1, Appli
817	31	57.4	760	3	US-08-767-820A-7	Sequence 7, Appli	c 890	31	57.4	1000	1	US-07-960-112B-3	Sequence 3, Appli
818	31	57.4	760	5	PCT-US95-06157-7	Sequence 7, Appli	c 891	31	57.4	1000	1	US-08-301-116B-1	Sequence 1, Appli
819	31	57.4	766	3	US-08-768-859A-9	Sequence 9, Appli	c 892	31	57.4	1000	1	US-08-301-116B-3	Sequence 3, Appli
820	31	57.4	766	3	US-08-767-820A-9	Sequence 9, Appli	c 893	31	57.4	1000	2	US-08-473-399B-1	Sequence 1, Appli
821	31	57.4	766	3	US-08-622-046B-6	Sequence 6, Appli	c 894	31	57.4	1000	2	US-08-473-399B-3	Sequence 3, Appli
822	31	57.4	766	3	US-08-622-046B-17	Sequence 17, Appli	c 895	31	57.4	1000	3	US-08-853-831-1	Sequence 1, Appli
823	31	57.4	766	3	US-09-100-264-6	Sequence 6, Appli	c 896	31	57.4	1000	3	US-08-853-831-3	Sequence 3, Appli
824	31	57.4	766	5	PCT-US95-06157-9	Sequence 9, Appli	c 897	31	57.4	1000	5	PCT-US93-09774-1	Sequence 3, Appli
825	31	57.4	772	4	US-09-020-956-11	Sequence 11, Appli	c 898	31	57.4	1000	5	PCT-US93-09774-3	Sequence 3, Appli
826	31	57.4	772	4	US-09-030-607-11	Sequence 11, Appli	c 899	31	57.4	1001	4	US-09-641-638-320	Sequence 320, App
827	31	57.4	772	4	US-09-605-785-11	Sequence 11, Appli	c 900	31	57.4	1002	4	US-09-328-111-159	Sequence 159, App
828	31	57.4	772	4	US-09-439-313-11	Sequence 11, Appli	c 901	31	57.4	1023	4	US-08-931-858E-49	Sequence 49, Appl
829	31	57.4	772	4	US-09-352-616A-11	Sequence 11, Appli	c 902	31	57.4	1032	4	US-09-897-537A-1	Sequence 1, Appli
830	31	57.4	772	4	US-09-232-149A-11	Sequence 11, Appli	c 903	31	57.4	1041	1	US-08-213-403-1	Sequence 1, Appli
831	31	57.4	779	4	US-08-894-251A-3	Sequence 3, Appli	c 904	31	57.4	1041	1	US-08-458-077-1	Sequence 1, Appli
832	31	57.4	779	4	US-09-020-956-10	Sequence 10, Appli	c 905	31	57.4	1041	1	US-08-460-741-1	Sequence 1, Appli
833	31	57.4	789	4	US-09-030-607-10	Sequence 10, Appli	c 906	31	57.4	1041	1	US-08-747-240-1	Sequence 1, Appli
834	31	57.4	789	4	US-09-605-785-10	Sequence 10, Appli	c 907	31	57.4	1055	3	US-09-215-131-3	Sequence 3, Appli
835	31	57.4	789	4	US-09-439-313-10	Sequence 10, Appli	c 908	31	57.4	1055	3	US-09-222-734-3	Sequence 3, Appli
836	31	57.4	789	4	US-09-352-616A-10	Sequence 10, Appli	c 909	31	57.4	1100	2	US-08-179-557-20	Sequence 20, Appl
837	31	57.4	789	4	US-09-232-149A-10	Sequence 10, Appli	c 910	31	57.4	1103	4	US-08-927-219-54	Sequence 54, Appl
838	31	57.4	801	4	US-09-161-241-3	Sequence 3, Appli	c 911	31	57.4	1126	1	US-08-233-788A-48	Sequence 48, Appl
839	31	57.4	822	4	US-09-100-264-8	Sequence 8, Appli	c 912	31	57.4	1157	4	US-08-934-386-5	Sequence 5, Appli
840	31	57.4	832	3	US-08-768-859A-5	Sequence 5, Appli	c 913	31	57.4	1165	4	US-08-448-722A-1	Sequence 1, Appli
841	31	57.4	832	3	US-08-768-859A-20	Sequence 20, Appli	c 914	31	57.4	1165	4	US-08-189-309B-1	Sequence 1, Appli
842	31	57.4	832	3	US-08-767-820A-5	Sequence 5, Appli	c 915	31	57.4	1170	3	US-08-689-421-20	Sequence 20, Appl
843	31	57.4	832	3	US-08-767-820A-20	Sequence 20, Appli	c 916	31	57.4	1170	4	US-09-389-528-20	Sequence 20, Appl
844	31	57.4	832	3	US-08-622-046B-4	Sequence 4, Appli	c 917	31	57.4	1170	4	US-09-181-827A-20	Sequence 20, Appl
845	31	57.4	832	3	US-08-622-046B-15	Sequence 15, Appli	c 918	31	57.4	1211	2	US-08-997-080-40	Sequence 40, Appl
846	31	57.4	832	5	PCT-US95-06157-5	Sequence 5, Appli	c 919	31	57.4	1211	2	US-08-997-362-40	Sequence 40, Appl
c 847	31	57.4	837	3	US-08-549-515-4	Sequence 4, Appli	c 920	31	57.4	1211	3	US-08-873-970-40	Sequence 40, Appl
848	31	57.4	838	1	US-08-633-779-1	Sequence 1, Appli	c 921	31	57.4	1211	4	US-09-095-855-40	Sequence 40, Appl
849	31	57.4	843	1	US-08-040-548-9	Sequence 9, Appli	c 922	31	57.4	1211	4	US-08-705-347A-40	Sequence 40, Appl
850	31	57.4	843	1	US-08-466-344-9	Sequence 9, Appli	c 923	31	57.4	1211	4	US-09-324-542-40	Sequence 40, Appl
c 851	31	57.4	856	2	US-09-057-762-23	Sequence 23, Appli	c 924	31	57.4	1211	4	US-09-205-426-40	Sequence 40, Appl
c 852	31	57.4	856	3	US-08-326-119A-23	Sequence 23, Appli	c 925	31	57.4	1211	4	US-09-200-643-40	Sequence 40, Appl
c 853	31	57.4	867	1	US-08-368-852-14	Sequence 14, Appli	c 926	31	57.4	1221	1	US-08-229-287-3	Sequence 3, Appli
c 854	31	57.4	867	2	US-08-525-940-14	Sequence 14, Appli	c 927	31	57.4	1231	1	US-08-197-792-40	Sequence 40, Appl
c 855	31	57.4	867	2	US-08-976-838-14	Sequence 14, Appli	c 928	31	57.4	1237	1	US-08-459-850-40	Sequence 40, Appl
c 856	31	57.4	893	1	US-08-333-321-1	Sequence 1, Appli	c 929	31	57.4	1237	1	US-08-459-214-40	Sequence 40, Appl
857	31	57.4	894	1	US-08-534-910B-1	Sequence 1, Appli	c 930	31	57.4	1263	2	US-08-828-488-6	Sequence 6, Appli
858	31	57.4	894	1	US-08-534-910B-2	Sequence 2, Appli	c 931	31	57.4	1263	4	US-09-299-689A-6	Sequence 6, Appli
859	31	57.4	894	1	US-08-534-910B-3	Sequence 3, Appli	c 932	31	57.4	1281	4	US-08-768-373-1	Sequence 1, Appli
860	31	57.4	894	1	US-08-534-910B-4	Sequence 4, Appli	c 933	31	57.4	1282	4	US-09-417-455-4	Sequence 4, Appli
861	31	57.4	894	1	US-08-534-910B-5	Sequence 5, Appli	c 934	31	57.4	1282	4	US-09-348-942-4	Sequence 4, Appli
862	31	57.4	894	3	US-08-886-466-1	Sequence 1, Appli	c 935	31	57.4	1282	4	US-09-457-626-4	Sequence 4, Appli
863	31	57.4	894	4	US-09-475-304-1	Sequence 4, Appli	c 936	31	57.4	1311	1	US-08-259-148A-9	Sequence 9, Appli
864	31	57.4	894	4	US-09-101-126-4	Sequence 4, Appli	c 937	31	57.4	1311	1	US-08-484-054-9	Sequence 9, Appli
865	31	57.4	894	4	US-09-367-528A-2	Sequence 2, Appli	c 938	31	57.4	1311	2	US-07-876-941A-9	Sequence 9, Appli
866	31	57.4	894	4	US-09-367-528A-4	Sequence 4, Appli	c 939	31	57.4	1311	4	US-07-870-985A-9	Sequence 9, Appli
c 867	31	57.4	897	2	US-09-006-535-2	Sequence 2, Appli	c 940	31	57.4	1313	2	US-08-463-911-6	Sequence 6, Appli
c 868	31	57.4	897	2	US-09-006-535-7	Sequence 7, Appli	c 941	31	57.4	1330	2	US-09-036-582-33	Sequence 33, Appli
c 869	31	57.4	920	4	US-09-221-017B-254	Sequence 254, App	c 942	31	57.4	1341	4	US-08-983-075D-6	Sequence 6, Appli
870	31	57.4	953	4	US-08-438-745-3	Sequence 3, Appli	c 943	31	57.4	1350	4	US-09-511-625B-60	Sequence 60, Appl
c 871	31	57.4	953	4	US-08-438-745-5	Sequence 5, Appli	c 944	31	57.4	1356	1	US-08-143-497-1	Sequence 1, Appli
c 872	31	57.4	953	4	US-09-219-019-3	Sequence 3, Appli	c 945	31	57.4	1356	1	US-08-461-666-1	Sequence 1, Appli
873	31	57.4	953	4	US-09-219-019-5	Sequence 5, Appli	c 946	31	57.4	1356	1	US-08-461-184-1	Sequence 1, Appli
874	31	57.4	953	5	PCT-US94-05669A-3	Sequence 3, Appli	c 947	31	57.4	1356	1	US-08-463-675-1	Sequence 1, Appli
875	31	57.4	953	5	PCT-US94-05669A-5	Sequence 5, Appli	c 948	31	57.4	1356	1	US-08-757-300-1	Sequence 1, Appli
c 876	31	57.4	984	1	US-08-240-049B-7	Sequence 7, Appli	c 949	31	57.4	1356	1	US-08-484-589-1	Sequence 1, Appli
c 877	31	57.4	984	1	US-08-259-148A-7	Sequence 7, Appli	c 950	31	57.4	1358	4	US-08-983-075D-8	Sequence 8, Appli
c 878	31	57.4	984	1	US-08-484-054-7	Sequence 7, Appli	c 951	31	57.4	1375	2	US-08-468-812-1	Sequence 1, Appli
c 879	31	57.4	984	2	US-07-876-941A-7	Sequence 7, Appli	c 952	31	57.4	1375	4	US-08-590-563-1	Sequence 1, Appli
c 880	31	57.4	984	4	US-08-542-634-5	Sequence 5, Appli	c 953	31	57.4	1376	2	US-08-868-288A-2	Sequence 2, Appli
c 881	31	57.4	984	4	US-08-477-292-5	Sequence 5, Appli	c 954	31	57.4	1376	3	US-09-235-373-2	Sequence 2, Appli
c 882	31	57.4	984	4	US-07-870-985A-7	Sequence 7, Appli	c 955	31	57.4	1430	2	US-09-388-993-2	Sequence 2, Appli
c 883	31	57.4	984	5	PCT-US95-13703-5	Sequence 5, Appli	c 956	31	57.4	1430	2	US-08-549-004A-15	Sequence 15, Appl
884	31	57.4	985	4	US-09-417-455-2	Sequence 2, Appli	c 957	31	57.4	1430	4	US-09-051-982A-15	Sequence 15, Appl
885	31	57.4	985	4	US-09-348-942-2	Sequence 2, Appli	c 958	31	57.4	1434	4	US-08-487-429A-6	Sequence 6, Appli
886	31	57.4	985	4	US-09-457-626-2	Sequence 2, Appli	c 959	31	57.4	1458	4	US-08-818-112-8	Sequence 8, Appli
887	31	57.4	986	4	US-09-501-192-8	Sequence 8, Appli	c 960	31	57.4	1458	4	US-08-818-111-8	Sequence 8, Appli
888	31	57.4	999	1	US-07-893-447-3	Sequence 3, Appli	c 961	31	57.4	1458	4	US-09-056-556-8	Sequence 8, Appli

962 31 57.4 1458 4 US-09-072-596-8 Sequence 8, Appli  
963 31 57.4 1497 1 US-08-488-961-5 Sequence 5, Appli  
964 31 57.4 1497 4 US-08-973-297-5 Sequence 5, Appli  
965 31 57.4 1497 5 PCT-US96-06511-5 Sequence 5, Appli  
966 31 57.4 1507 4 US-08-969-815-1 Sequence 1, Appli  
967 31 57.4 1507 4 US-09-120-025-1 Sequence 1, Appli  
968 31 57.4 1507 4 US-09-710-481-1 Sequence 1, Appli  
969 31 57.4 1525 2 US-09-006-535-1 Sequence 1, Appli  
970 31 57.4 1525 6 5292979-1 Patent No. 5292979  
971 31 57.4 1529 4 US-09-149-476-250 Sequence 250, App  
972 31 57.4 1551 2 US-08-828-488-4 Sequence 4, Appli  
973 31 57.4 1551 4 US-09-299-689A-4 Sequence 4, Appli  
974 31 57.4 1619 3 US-09-163-162-1 Sequence 1, Appli  
975 31 57.4 1619 4 US-09-286-407-1 Sequence 1, Appli  
976 31 57.4 1619 4 US-09-496-694B-97 Sequence 97, Appli  
977 31 57.4 1647 4 US-08-542-634-3 Sequence 3, Appli  
978 31 57.4 1647 4 US-08-477-292-3 Sequence 3, Appli  
979 31 57.4 1647 5 PCT-US95-13703-3 Sequence 3, Appli  
980 31 57.4 1648 4 US-09-202-548B-11 Sequence 11, Appli  
981 31 57.4 1670 2 US-08-828-488-2 Sequence 2, Appli  
982 31 57.4 1670 4 US-09-299-689A-2 Sequence 2, Appli  
983 31 57.4 1670 4 US-09-071-035-267 Sequence 267, App  
984 31 57.4 1720 3 US-08-555-352-11 Sequence 11, Appli  
985 31 57.4 1720 4 US-09-258-016-11 Sequence 11, Appli  
986 31 57.4 1720 4 US-09-257-825B-11 Sequence 11, Appli  
987 31 57.4 1726 5 PCT-US92-02091-3 Sequence 3, Appli  
988 31 57.4 1770 2 US-08-317-305-1 Sequence 1, Appli  
989 31 57.4 1770 3 US-08-862-508-1 Sequence 1, Appli  
990 31 57.4 1770 5 PCT-US95-12508-1 Sequence 1, Appli  
991 31 57.4 1778 4 US-08-934-386-4 Sequence 4, Appli  
992 31 57.4 1830 4 US-09-149-476-129 Sequence 129, App  
993 31 57.4 1855 4 US-08-810-009-1 Sequence 1, Appli  
994 31 57.4 1860 4 US-09-178-252-19 Sequence 19, Appli  
995 31 57.4 1886 1 US-07-980-526-1 Sequence 1, Appli  
996 31 57.4 1887 4 US-09-267-311-3 Sequence 3, Appli  
997 31 57.4 1907 3 US-08-771-986A-1 Sequence 1, Appli  
998 31 57.4 1907 3 US-08-771-986A-3 Sequence 3, Appli  
999 31 57.4 1907 3 US-08-769-802A-1 Sequence 1, Appli  
1000 31 57.4 1978 4 US-09-198-603C-5 Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-09-221-017B-726  
: Sequence 726, Application US/09221017B  
: Patent No. 644799  
: GENERAL INFORMATION:  
: APPLICANT: Ross, Bruce C.  
: TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
: NUMBER OF SEQUENCES: 1120  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: MORRISON & FOERSTER  
: STREET: 755 PAGE MILL ROAD  
: CITY: Palo Alto  
: STATE: CA  
: COUNTRY: USA  
: ZIP: 94304-1018  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: Windows  
: SOFTWARE: FASTSQ for Windows Version 2.0b  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/221,017B  
: FILING DATE: 23-DEC-1998  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PPI182  
: FILING DATE: 31-DEC-1997  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PPI546  
: FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: PP2911  
: FILING DATE: 09-APR-1998  
: PRIOR APPLICATION DATA: PCT/AU98/01023  
: APPLICATION NUMBER: PCT/AU98/01023  
: FILING DATE: 10-DEC-1998  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Monroy, Gladys H  
: REGISTRATION NUMBER: 32,430  
: REFERENCE/DOCKET NUMBER: 27340-20021.00  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 650-813-5600  
: TELEFAX: 650-494-0792  
: TELEX: 706141  
: INFORMATION FOR SEQ ID NO: 726:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1974 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: circular  
: MOLECULE TYPE: DNA (genomic)  
: HYPOTHEetical: NO  
: ANTI-SENSE: UNKNOWN  
: ORIGINAL SOURCE:  
: ORGANISM: PORYPHYROMONAS GINGIVALIS  
: FEATURE:  
: NAME/KEY: misc.feature  
: LOCATION: 1...1974  
US-09-221-017B-726  
Alignment Scores: 0.605 Length: 1974  
Pred. No.: 54.00 Matches: 10  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 100.00% Gaps: 0  
DB: 4  
US-10-008-355-25 (1-10) x US-09-221-017B-726 (1-1974)  
Oy 1 ThrGlyGlyAsnSerGlySerProValphe 10  
Db 1524 ACGGGCGGTACTCCGGTAGCCCGGTATTC 1553  
RESULT 2  
US-09-134-001C-892  
: Sequence 892, Application US/09134001C  
: Patent No. 6380370  
: GENERAL INFORMATION:  
: APPLICANT: Lynn Doucette-Stamm et al  
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
: CORRESPONDENCE ADDRESS: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
: FILE REFERENCE: GTC-007  
: CURRENT APPLICATION NUMBER: US/09/134,001C  
: CURRENT FILING DATE: 1998-08-13  
: PRIOR APPLICATION NUMBER: US 60/064,964  
: PRIOR FILING DATE: 1997-11-08  
: PRIOR APPLICATION NUMBER: US 60/055,779  
: PRIOR FILING DATE: 1997-08-14  
: NUMBER OF SEQ ID NOS: 5674  
: SEQ ID NO 892  
: LENGTH: 936  
: TYPE: DNA  
: ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-892  
Alignment Scores: 2.17 Length: 936  
Pred. No.: 49.00 Matches: 9  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 90.74% Gaps: 0  
DB: 4



FILING DATE: 24-OCT-1991  
APPLICATION NUMBER: Japanese Patent  
APPLICATION NUMBER: Application No. 5459064 2-288110  
FILING DATE: October 24, 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: BOZICEVIC, KARL.  
REGISTRATION NUMBER: 28,807  
REFERENCE/DOCKET NUMBER: 29900-20298.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 677-7000  
TELEFAX: (415) 677-7522  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1448 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Bacillus licheniformis  
STRAIN: ATCC NO. 14580  
FEATURE:  
NAME/KEY: coding sequence  
LOCATION: 323 to 1270  
IDENTIFICATION METHOD: by experiment  
NAME/KEY: signal peptide  
LOCATION: 323 to 604  
IDENTIFICATION METHOD: by experiment  
NAME/KEY: mature peptide  
LOCATION: 605 to 1270  
IDENTIFICATION METHOD: by experiment  
OTHER INFORMATION:  
OTHER INFORMATION: Xaa at -94 position of amino acid  
OTHER INFORMATION: sequence: formyl methionine  
US-08-035-634-1

Alignment Scores: 44.7 Length: 1448  
Prd. No.: 43.00 Matches: 8  
Score: 88.89% Conservative: 0  
Percent Similarity: 88.89% Mismatches: 1  
Best Local Similarity: 88.89% Indels: 0  
Query Match: 79.63% Gaps: 0  
DB: 1

US-10-008-355-25 (1-10) x US-08-035-634-1 (1-1448)

QY 2 GlyGlyAsnSerGlySerProValPhe 10  
||||| |||||||||  
DB 1094 GGAGACAAAGCGGTTCCCGGTATTC 1120

RESULT 7  
US-08-998-416-856/c  
; Sequence 856, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Rebschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 856:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 743 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: PAG15390P  
US-08-998-416-856

Alignment Scores:  
Pred. No.: 75.3 Length: 743  
Score: 40.00 Matches: 6  
Percent Similarity: 80.00% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 2  
Query Match: 74.07% Indels: 0  
DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-08-998-416-856 (1-743)  
QY 1 Thrglyglynserserglyserprovalphe 10  
Db 576 ACCGGTGGATCTGGAGACCTATATAT 547

RESULT 8  
US-08-626-169-2  
Sequence 2, Application US/08626169  
Patent No. 5861248  
GENERAL INFORMATION:  
APPLICANT: Russell, David W.  
APPLICANT: Thigpen, Anice E.  
TITLE OF INVENTION: BIONARKERS FOR DETECTION, DIAGNOSIS  
TITLE OF INVENTION: AND PROGNOSIS OF PROSTATE CANCER  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
City: Houston  
STATE: Texas  
COUNTRY: United States  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/626,169  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Corder, Timothy S.  
REGISTRATION NUMBER: 38,414  
REFERENCE/DOCKET NUMBER: UROC.007

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-626-169-2

Alignment Scores:  
Pred. No.: 32.5 Length: 240  
Score: 39.00 Matches: 7  
Percent Similarity: 77.78% Conservative: 0  
Best Local Similarity: 77.78% Mismatches: 2  
Query Match: 72.22% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x US-08-626-169-2 (1-240)  
QY 2 glyglynserserglyserprovalphe 10  
Db 6 GGGGGGAACGTGGGGAACCTGTGTTT 32

RESULT 9  
US-09-164-907-2  
Sequence 2, Application US/09164907A  
Patent No. 6090559  
GENERAL INFORMATION:  
APPLICANT: RUSSELL, DAVID W.  
APPLICANT: THIGPEN, ANICE E.  
TITLE OF INVENTION: BIONARKERS FOR DETECTION, DIAGNOSIS AND PROGNOSIS OF  
TITLE OF INVENTION: PROSTATE CANCER  
FILE REFERENCE: UROC.021  
CURRENT APPLICATION NUMBER: US/09/164,907A  
CURRENT FILING DATE: 1998-10-01  
EARLIER APPLICATION NUMBER: 08/626,169  
EARLIER FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 240  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-164-907-2

Alignment Scores:  
Pred. No.: 32.5 Length: 240  
Score: 39.00 Matches: 7  
Percent Similarity: 77.78% Conservative: 0  
Best Local Similarity: 77.78% Mismatches: 2  
Query Match: 72.22% Indels: 0  
DB: 3 Gaps: 0

US-10-008-355-25 (1-10) x US-09-164-907-2 (1-240)  
QY 2 glyglynserserglyserprovalphe 10  
Db 6 GGGGGGAACGTGGGGAACCTGTGTTT 32

RESULT 10  
US-07-723-002C-5  
Sequence 5, Application US/07723002C  
Patent No. 5447862  
GENERAL INFORMATION:  
APPLICANT: Heim, Jutta  
APPLICANT: Meyhack, Bernd  
APPLICANT: Gysler, Christof  
APPLICANT: Visser, Jacob  
APPLICANT: Kester, Hermanus Cornelis Maria  
TITLE OF INVENTION: No. 5447862el Expression System  
NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/723,002C  
 FILING DATE: 28-JUN-1991  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 8702475  
 FILING DATE: 04-FEB-1987  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 150,880  
 FILING DATE: 29-JAN-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 8818046.8  
 FILING DATE: 28-JUL-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 8914666.6  
 FILING DATE: 26-JUN-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 384,898  
 FILING DATE: 24-JUL-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Foley, Shawn P.  
 REGISTRATION NUMBER: 33,071  
 REFERENCE/DOCKET NUMBER: 4-16317/+/CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8615  
 TELEFAX: (919)541-8689  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2774 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Aspergillus niger N400  
 IMMEDIATE SOURCE:  
 CLONE: Plasmid pgw830 (DSM 4389)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: join(1134..1337, 1400..1543, 1600..1725, 1783  
 LOCATION: ..2112, 2170..2502)  
 OTHER INFORMATION: /transl\_except= (pos: 1541 .. 1543, aa: Tyr)  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 1134..1193  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1194..1337  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1338..1399  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1400..1542  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1543..1598  
 FEATURE:  
 NAME/KEY: exon

LOCATION: 1599..1725  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1726..1782  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1783..2112  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 2113..2169  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2170..2502  
 FEATURE:  
 NAME/KEY: promoter  
 LOCATION: 1..1133  
 FEATURE:  
 NAME/KEY: terminator  
 LOCATION: 2506..2774  
 US-07-723-002C-5  
 Alignment Scores:  
 Pred. No.: 500  
 Score: 39.00  
 Percent Similarity: 80.00%  
 Best Local Similarity: 70.00%  
 Query Match: 72.22%  
 DB: 1  
 Gaps: 0  
 Length: 2774  
 Matches: 7  
 Conservative: 1  
 Mismatches: 2  
 Indels: 0  
 Gaps: 0  
 US-10-008-355-25 (1-10) x US-07-723-002C-5 (1-2774)  
 QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
 Db 1236 ACTGGCGGTGGCAGCGCTTCCCGCTCTAT 1265  
 |||||  
 US-09-222-575-123  
 ; Sequence 123, Application US/09222575  
 ; Patent No. 6387697  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yugitu, Jiang  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Xu, Jiangchun  
 ; TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer  
 ; FILE REFERENCE: 210121.470  
 ; CURRENT APPLICATION NUMBER: US/09/222,575  
 ; CURRENT FILING DATE: 1998-12-28  
 ; NUMBER OF SEQ ID NOS: 174  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 123  
 ; LENGTH: 531  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: modified\_base  
 ; LOCATION: (152)  
 ; OTHER INFORMATION: where n is a, c, g or t  
 ; NAME/KEY: modified\_base  
 ; LOCATION: (373)  
 ; OTHER INFORMATION: where n is a, c, g or t  
 ; NAME/KEY: modified\_base  
 ; LOCATION: (482)  
 ; OTHER INFORMATION: where n is a, c, g or t  
 ; NAME/KEY: modified\_base  
 ; LOCATION: (494)  
 ; OTHER INFORMATION: where n is a, c, g or t  
 ; NAME/KEY: modified\_base  
 ; LOCATION: (496)  
 ; OTHER INFORMATION: where n is a, c, g or t  
 ; NAME/KEY: modified\_base  
 ; LOCATION: (502)

OTHER INFORMATION: where n is a, c, g or t  
US-09-222-575-123  
Alignment Scores:  
Pred. No.: 120 Length: 531  
Score: 38.00 Matches: 6  
Percent Similarity: 77.78% Conservative: 1  
Best Local Similarity: 66.57% Mismatches: 2  
Query Match: 70.37% Indels: 0  
DB: 4 Gaps: 0  
US-10-008-355-25 (1-10) x US-09-222-575-123 (1-531)  
QY 2 gLyGlyAsnSerGlySerProValPhe 10  
||||| |||||||:  
Db 241 GGGGAGAGTGGGGAAGCCCTATTTTT 267  
RESULT 12  
US-07-624-313-3/c  
Sequence 3, Application US/07624313  
Patent No. 5250411  
GENERAL INFORMATION:  
APPLICANT: Ayanathan, K.  
APPLICANT: Bhat, P.  
APPLICANT: Datta, S.  
APPLICANT: Francis, V.S.N.K.  
APPLICANT: Padmanaban, G.  
APPLICANT: Srinivasa, H.  
TITLE OF INVENTION: NEW ANALYSIS METHOD  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/624,313  
FILING DATE: 19901204  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 8904100-8  
FILING DATE: 05-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Ryan, John W.  
REGISTRATION NUMBER: 33,771  
REFERENCE/DOCKET NUMBER: 1103326-811  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-819-8515  
TELEFAX: 212-354-8113  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 663 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium vivax  
IMMEDIATE SOURCE:  
CLONE: PARC 1153  
US-07-624-313-3  
Alignment Scores: 154 Length: 663  
Pred. No.: 154 Length: 663

Score: 38.00 Matches: 7  
Percent Similarity: 77.78% Conservative: 0  
Best Local Similarity: 77.78% Mismatches: 2  
Query Match: 70.37% Indels: 0  
DB: 1 Gaps: 0  
US-10-008-355-25 (1-10) x US-07-624-313-3 (1-663)  
QY 2 gLyGlyAsnSerGlySerProValPhe 10  
||||| |||||||:  
Db 220 GGAGGAACAGTGGGAACCTGGATT 194  
RESULT 13  
US-09-061-709-8  
Sequence 8, Application US/09061709B  
Patent No. 6297364  
GENERAL INFORMATION:  
APPLICANT: Chen, Yao-Tseng  
APPLICANT: Gure, Ali  
APPLICANT: Tsang, Solam  
APPLICANT: Stockert, Elisabeth  
APPLICANT: Jager, Elke  
APPLICANT: Knuth, Alexander  
APPLICANT: Old, Lloyd J.  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated  
FILE REFERENCE: LUD 5538  
CURRENT APPLICATION NUMBER: US/09/061,709B  
CURRENT FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 8  
SEQ ID NO 8  
LENGTH: 3283  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
US-09-061-709-8  
Alignment Scores:  
Pred. No.: 922 Length: 3283  
Score: 38.00 Matches: 7  
Percent Similarity: 90.00% Conservative: 2  
Best Local Similarity: 70.00% Mismatches: 1  
Query Match: 70.37% Indels: 0  
DB: 4 Gaps: 0  
US-10-008-355-25 (1-10) x US-09-061-709-8 (1-3283)  
QY 1 ThGlyGlyAsnSerGlySerProValPhe 10  
||| |||:  
Db 2406 ACTCAGCAGGTCTGTGATCAGCTGTAT 2435  
RESULT 14  
US-09-061-709-6  
Sequence 6, Application US/09061709B  
Patent No. 6297364  
GENERAL INFORMATION:  
APPLICANT: Chen, Yao-Tseng  
APPLICANT: Gure, Ali  
APPLICANT: Tsang, Solam  
APPLICANT: Stockert, Elisabeth  
APPLICANT: Jager, Elke  
APPLICANT: Knuth, Alexander  
APPLICANT: Old, Lloyd J.  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated  
FILE REFERENCE: LUD 5538  
CURRENT APPLICATION NUMBER: US/09/061,709B  
CURRENT FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 8  
SEQ ID NO 6  
LENGTH: 3412  
TYPE: DNA  
ORGANISM: Homo sapiens

```

FEATURE:
US-09-061-709-6

Alignment Scores:
Pred. No.: 962 Length: 3412
Percent Score: 38.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 70.37% Indels: 0
DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-061-709-6 (1-3412)
Oy 1 ThrGlyGlyAnSerGlySerProValphe 10
||| |||:|||||:|||||:|||||:
Db 2535 ACTCAGCGCAGTTCGTGGATCACCCTGTGTAT 2564

RESULT 15
US-08-225-224-31/c
Sequence 31, Application US/08225224
Patent No. 5635599
GENERAL INFORMATION:
APPLICANT: PASTIAN, Ira
APPLICANT: KEITMAN, Robert J.
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend Kourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,224
FILING DATE: 8-APR-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen L.
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 15280-193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 base pairs
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: DNA (primer)
US-08-225-224-31

Alignment Scores:
Pred. No.: 9.93 Length: 39
Score: 37.00 Matches: 6
Percent Similarity: 87.50% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 68.52% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x US-08-225-224-31 (1-39)
Oy 1 ThrGlyGlyAnSerGlySerPro 8
||||| |||:||||| |||:|||||
Db 24 ACCGAGAGTAAAGGTGGGCGACCT 1

```

```

RESULT 16
US-08-722-258-31/c
; Sequence 31, Application US/08722258
; Patent No. 6011002
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Kreltman, Robert J.
; APPLICANT: Puri, Raj K.
; TITLE OF INVENTION: Circularly Permuted Ligands and
; TITLE OF INVENTION: Circularly Permuted Chimeric Molecules
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,258
; FILING DATE: 08-JAN-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/04468
; FILING DATE: 06-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/225,224
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-193100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..39
; OTHER INFORMATION: /note= "BK-138 primer"
US-08-722-258-31

Alignment Scores:
Pred. No.:          9..93          Length:          39
Score:              37..00          Matches:         6
Percent Similarity: 87.50%          Conservative:    1
Best Local Similarity: 75.00%       Mismatches:     1
Query Match:        68..52%         Indels:         0
DB:                 3              Gaps:            0

US-10-008-355-25 (1-10) x US-08-722-258-31 (1-39)
QY      1  ThrGlyGlyAsnSerGlySerPro 8
      |||||
DB      24  ACCGGAGGTACACGGTGGGGCCACT 1

RESULT 17
PCT-US95-04468-31/c
; Sequence 31, Application PC/TUS9504468
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND

```



TITLE OF INVENTION: CIRCULARLY PERMUTED FUSION PROTEINS  
NUMBER OF SEQUENCES: 59  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04468  
FILING DATE: 07-APR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/225,224  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen L.  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 15280-193-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ. ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (primer)  
PCT-US95-04468-31

Alignment Scores:  
Pred. No.: 9.93 Length: 39  
Score: 37.00 Matches: 6  
Percent Similarity: 87.50% Conservative: 1  
Best Local Similarity: 75.00% Mismatches: 1  
Query Match: 68.52% Indels: 0  
DB: 5 Gaps: 0

US-10-008-355-25 (1-10) x PCT-US95-04468-31 (1-39)

QY 1 ThrglyGlyAsnSerGlySerPro 8  
|||||  
Db 24 ACCGAGTAACGCTGGGCACCT 1

RESULT 18  
US-08-225-224-28  
Sequence 28, Application US/08225224  
Patent No. 5635599  
GENERAL INFORMATION:  
APPLICANT: PASTAN, Ira  
APPLICANT: KREITMAN, Robert J.  
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND  
TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Stewart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,224  
FILING DATE: 8-APR-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen L.  
REGISTRATION NUMBER: 32,762

REFERENCE/DOCKET NUMBER: 15280-193  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ. ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (primer)  
US-08-225-224-28

Alignment Scores:  
Pred. No.: 10.8 Length: 42  
Score: 37.00 Matches: 6  
Percent Similarity: 87.50% Conservative: 1  
Best Local Similarity: 75.00% Mismatches: 1  
Query Match: 68.52% Indels: 0  
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x US-08-225-224-28 (1-42)

QY 1 ThrglyGlyAsnSerGlySerPro 8  
|||||  
Db 7 ACCGAGTAACGCTGGGCACCT 30

RESULT 19  
US-08-722-258-28  
Sequence 28, Application US/08722258  
Patent No. 6011002  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira  
APPLICANT: Kreitman, Robert J.  
APPLICANT: Puti, Raj K.  
TITLE OF INVENTION: Circularly Permuted Ligands and  
TITLE OF INVENTION: Circularly Permuted Chimeric Molecules  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,258  
FILING DATE: 08-JAN-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/04468  
FILING DATE: 06-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/225,224  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen Lauver  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 015280-193100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ. ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..42  
OTHER INFORMATION: /note= "BK-135 primer"  
US-08-722-258-28  
Alignment Scores:  
Pred. No.: 10.8 Length: 42  
Score: 37.00 Matches: 6  
Percent Similarity: 87.50% Conservative: 1  
Best Local Similarity: 75.00% Mismatches: 1  
Query Match: 68.52% Indels: 0  
DB: 3 Gaps: 0  
US-10-008-355-25 (1-10) x US-08-722-258-28 (1-42)  
QY 1 ThrglyglaSerglySerPro 8  
|||||  
Db 7 ACCGAGGTACGCTGGGGCACCT 30  
RESULT 20  
PCT-US95-04468-28  
Sequence 28, Application PC/TUS9504468  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND  
TITLE OF INVENTION: CIRCULARLY PERMUTED FUSION PROTEINS  
NUMBER OF SEQUENCES: 59  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04468  
FILING DATE: 07-APR-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/225,224  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen L.  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 15280-193-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (primer)  
PCT-US95-04468-28  
Alignment Scores:  
Pred. No.: 10.8 Length: 42  
Score: 37.00 Matches: 6  
Percent Similarity: 87.50% Conservative: 1  
Best Local Similarity: 75.00% Mismatches: 1  
Query Match: 68.52% Indels: 0  
DB: 5 Gaps: 0  
US-10-008-355-25 (1-10) x PCT-US95-04468-28 (1-42)  
QY 1 ThrglyglaSerglySerPro 8  
|||||  
Db 7 ACCGAGGTACGCTGGGGCACCT 30  
RESULT 21  
US-08-664-596B-21/c

Sequence 21, Application US/08664596B  
Patent No. 5807703  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Lavallee, Edward  
APPLICANT: Racie, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Evans, Cheryl  
APPLICANT: Spaulding, Vikki  
APPLICANT: Bowman, Michael  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/664,596B  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 633 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-664-596B-21  
Alignment Scores:  
Pred. No.: 224 Length: 633  
Score: 37.00 Matches: 6  
Percent Similarity: 88.89% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 1  
Query Match: 68.52% Indels: 0  
DB: 1 Gaps: 0  
US-10-008-355-25 (1-10) x US-08-664-596B-21 (1-633)  
QY 2 GlyglyaSerGlySerProValphe 10  
|||||  
Db 488 GTGGCACTCCAGCAATCCCTCTTC 462  
RESULT 22  
US-09-499-884-11  
Sequence 11, Application US/09499884  
Patent No. 6265172  
GENERAL INFORMATION:  
APPLICANT: St. Clair, Darci  
APPLICANT: Ureno, Muneyasu  
APPLICANT: Kasarskis, Edward  
TITLE OF INVENTION: DIAGNOSTIC TEST AND THERAPY FOR MANGANESE SUPEROXIDE DISMUTASE  
FILE REFERENCE: 50229-180  
CURRENT APPLICATION NUMBER: US/09/499,884  
CURRENT FILING DATE: 2000-02-08

NUMBER OF SEQ ID NOS: 12  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 11  
LENGTH: 3663  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-499-884-11

Alignment Scores:  
Pred. No.: 1.59e+03 Length: 3663  
Score: 37.00 Matches: 7  
Percent Similarity: 88.89% Conservative: 1  
Best Local Similarity: 77.78% Mismatches: 1  
Query Match: 68.52% Indels: 0  
DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-499-884-11 (1-3663)

Oy 2 GlycylAsnSerglySerProValphe 10  
|||||:||||| 111  
Db 331 GGGGAGACAGTGAGGAGCCAGACTTT 357

RESULT 23  
US-08-887-365-32  
Sequence 32, Application US/08887365  
Patent No. 5858760  
GENERAL INFORMATION:  
APPLICANT: Dalboege, Henrik  
APPLICANT: Kofod, Lene V.  
APPLICANT: Kauppinen, Markus S.  
APPLICANT: Andersen, Lene N.  
APPLICANT: Christgau, Stephan  
APPLICANT: Heldt-Hansen, Hans P.  
TITLE OF INVENTION: AN ENZYME WITH PECTIN LYASE ACTIVITY  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: No. 58587600 No. 5858760disk of No. 5858760th America, Inc.  
STREET: 405 Lexington Avenue, 64th floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/887,365.  
FILING DATE: 02-JUL-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/513,928  
FILING DATE: 26-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Harrington, James J.  
REGISTRATION NUMBER: 38,711  
REFERENCE/DOCKET NUMBER: 3955,204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9635  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 184 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-887-365-32

Alignment Scores:  
Pred. No.: 85.8 Length: 184  
Score: 36.00 Matches: 6  
Percent Similarity: 80.00% Conservative: 2

Best Local Similarity: 60.00% Mismatches: 2  
Query Match: 66.67% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x US-08-887-365-32 (1-184)

Oy 1 ThrGlyGlyAsnSerglySerProValphe 10  
||||| 111 :|||:|  
Db 113 ACTGCTGCTGCTGAGTGCAGACTCGGTTTAT 142

RESULT 24  
US-08-680-326-118/c  
Sequence 118, Application US/08680326  
Patent No. 5925733  
GENERAL INFORMATION:  
APPLICANT: ROSE, TIMOTHY M.  
APPLICANT: BOSCH, MARINX  
APPLICANT: STRAND, KURT  
APPLICANT: TODARO, GEORGE J.  
TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES  
TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL  
NUMBER OF SEQUENCES: 152  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,326  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Schiff, J. Michael  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 454 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..454  
US-08-680-326-118

Alignment Scores:  
Pred. No.: 235 Length: 454  
Score: 36.00 Matches: 6  
Percent Similarity: 88.89% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 1  
Query Match: 66.67% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x US-08-680-326-118 (1-454)

Oy 1 ThrGlyGlyAsnSerglySerProVal 9  
|||||:||||| 111111  
Db 295 ACCGAGGCGACGCCGGAACCGCA 269

RESULT 25

US-09-221-017B-38/C  
; Sequence 38, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FASTSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/221,017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: P01182  
; FILING DATE: 31-DEC-1997  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: P01546  
; FILING DATE: 30-JAN-1998  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: P02911  
; FILING DATE: 09-APR-1998  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: P01023  
; FILING DATE: 10-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MONROY, Gladys H  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-813-5600  
; TELEFAX: 650-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 686 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: UNKNOWN  
; ORIGINAL SOURCE:  
; ORGANISM: PORPHYROMONAS GINGIVALIS  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1...686  
; DB: 1...686  
US-09-221-017B-38

Alignment Scores:  
Pred. No.: 373 Length: 686  
Score: 36.00 Matches: 6  
Percent Similarity: 88.89% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 1  
Query Match: 66.67% Indels: 0  
Gaps: 0  
DB: 4

US-10-008-355-25 (1-10) x US-09-221-017B-38 (1-686)

OY 1 ThrGlyGlyAsnSerGlySerProVal 9  
|||||  
Db 482 ACGGGCGATCATCCGCTTCCCGCATC 456

RESULT 26  
US-08-887-365-35  
; Sequence 35, Application US/08887365  
; Patent No. 5858760  
; GENERAL INFORMATION:  
; APPLICANT: Dalboege, Henrik  
; APPLICANT: Kolof, Lene V.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Christgau, Stephan  
; APPLICANT: Heldt-Hansen, Hans P.  
; TITLE OF INVENTION: AN ENZYME WITH PECTIN LYASE ACTIVITY  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5858760 No. 5858760disk of No. 5858760th America, Inc.  
; STREET: 405 Lexington Avenue, 64th floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/887,365  
; FILING DATE: 02-JUL-1997  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/513,928  
; FILING DATE: 26-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J.  
; REGISTRATION NUMBER: 38,711  
; REFERENCE/DOCKET NUMBER: 3955.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1291 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 15..1151  
; DB: 15..1151  
US-08-887-365-35

Alignment Scores:  
Pred. No.: 756 Length: 1291  
Score: 36.00 Matches: 6  
Percent Similarity: 80.00% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 2  
Query Match: 66.67% Indels: 0  
Gaps: 0  
DB: 2

US-10-008-355-25 (1-10) x US-08-887-365-35 (1-1291)

OY 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
|||||  
Db 114 ACTGTTGCTGAGTGCAGACTCCGCTTAT 143

RESULT 27  
US-07-952-853-21  
; Sequence 21, Application US/07952853  
; Patent No. 5863783  
; GENERAL INFORMATION:  
; APPLICANT: Van Heuvel, Margaretha  
; APPLICANT: Bakhuys, Janna G.  
; APPLICANT: Couetil, Yves  
; APPLICANT: Harder, Abraham

APPLICANT: De Graaff, leendert H.  
 APPLICANT: Filippi, Michel J. A.  
 APPLICANT: Van Der Veen, Peter  
 APPLICANT: Visser, Jacob  
 APPLICANT: Andreoli, Peter M.  
 TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA  
 TITLE OF INVENTION: MOLECULES  
 TITLE OF INVENTION: ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL  
 TITLE OF INVENTION: ORIGIN  
 NUMBER OF SEQUENCES: 24  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MORRISON & FOERSTER  
 STREET: 755 Page Mill Road  
 CITY: Palo Alto  
 STATE: California  
 ZIP: 94304-1018  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/952,853  
 FILING DATE: 19921125  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murashige, Kate H.  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 246152003500  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-813-5600  
 TELEFAX: 415-494-0792  
 TELEX: 706141  
 INFORMATION FOR SEQ. ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3938 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Aspergillus niger  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1247..1390  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1391..1442  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1443..1957  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 1958..2005  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2006..2089  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 2090..2137  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2138..2214  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 2215..2262  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2263..2295  
 FEATURE:  
 NAME/KEY: Intron

LOCATION: 2296..2346  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2347..2498  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 2499..2548  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2549..3037  
 FEATURE:  
 NAME/KEY: Intron  
 LOCATION: 3038..3092  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 3093..3485  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: join(1247..1390, 1443..1957, 2006..2089,  
 LOCATION: 2138..2214, 2263..2295, 2347..2498, 2549..3037,  
 LOCATION: 3093..3485)  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start=1247  
 OTHER INFORMATION: /product="alpha-L-arabinofuranosidase A"  
 OTHER INFORMATION: /evidence=EXPERIMENTAL  
 OTHER INFORMATION: /gene="exonA"  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 1247..1321  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 LOCATION: 1322..3485  
 US-07-952-853-21  
 Alignment Scores:  
 Pred. No.: 2.64e+03 Length: 3938  
 Score: 36.00 Matches: 6  
 Percent Similarity: 87.50% Conservative: 1  
 Best Local Similarity: 75.00% Mismatches: 1  
 Query Match: 66.67% Indels: 0  
 DB: 2 Gaps: 0  
 US-10-008-355-25 (1-10) x US-07-952-853-21 (1-3958)  
 Qy 2 GLYGLYASNSERGLYSERPROVAL 9  
 Db 1346 GGTGGCACTCATCCAGCCCATC 1369  
 RESULT 28  
 US-08-914-848-21  
 Sequence 21. Application US/08914848  
 Patent No. 5983887  
 GENERAL INFORMATION:  
 APPLICANT: Van Heuvel, Margaretha  
 APPLICANT: Bakhuys, Janna G.  
 APPLICANT: Coustel, Yves  
 APPLICANT: Harder, Abraham  
 APPLICANT: De Graaff, leendert H.  
 APPLICANT: Filippi, Michel J. A.  
 APPLICANT: Van Der Veen, Peter  
 APPLICANT: Visser, Jacob  
 APPLICANT: Andreoli, Peter M.  
 TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA  
 TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL  
 TITLE OF INVENTION: ORIGIN  
 NUMBER OF SEQUENCES: 24  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MORRISON & FOERSTER  
 STREET: 755 Page Mill Road  
 CITY: Palo Alto  
 STATE: California  
 ZIP: 94304-1018  
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/914,848  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/952,853  
 FILING DATE: 25-NOV-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Murashige, Kate H.  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 246152003500  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-813-5600  
 TELEFAX: 415-494-0792  
 TELEX: 706141  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3958 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: *Aspergillus niger*  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1247..1390  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1391..1442  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 1443..1957  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1958..2005  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2006..2089  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 2090..2137  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2138..2214  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 2215..2262  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2263..2295  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 2296..2346  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2347..2498  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 2499..2548  
 FEATURE:  
 NAME/KEY: exon  
 LOCATION: 2549..3037  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 3038..3092  
 FEATURE:

NAME/KEY: exon  
 LOCATION: 3093..3485  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: join(1247..1390, 1443..1957, 2006..2089, 2138..2214, 2263..2295, 2347..2498, 2549..3037, 3093..3485)  
 IDENTIFICATION METHOD: experimental  
 OTHER INFORMATION: /codon\_start=1247  
 OTHER INFORMATION: /product="alpha-L-arabinofuranosidase A"  
 OTHER INFORMATION: /evidence=EXPERIMENTAL  
 OTHER INFORMATION: /gene="exonA"  
 FEATURE:  
 NAME/KEY: sig\_peptide  
 LOCATION: 1247..1321  
 FEATURE:  
 NAME/KEY: mat\_peptide  
 LOCATION: 1322..3485  
 US-08-914-848-21  
 Alignment Scores:  
 Pred. No.: 2.64e+03 Length: 3958  
 Score: 36.00 Matches: 6  
 Percent Similarity: 87.50% Conservative: 1  
 Best Local Similarity: 75.00% Mismatches: 1  
 Query Match: 66.67% Indels: 0  
 DB: 2 Gaps: 0  
 US-10-008-355-25 (1-10) x US-08-914-848-21 (1-3958)  
 QY 2 Glycylasnserylglyserproval 9  
 Db 1346 GGTGGCACTCATCCAGCCCATC 1369  
 RESULT 29  
 US-09-390-721-1  
 ? Sequence 1, Application US/09390721  
 ? Patent No. 6197591  
 ? GENERAL INFORMATION:  
 ? APPLICANT: STUTZMAN-ENGWALL, KIM J.  
 ? TITLE OF INVENTION: STREPTOMYCES AVERMITILIS REGULATORY GENES FOR INCREASED  
 ? FILE REFERENCE: PC9944A  
 ? CURRENT APPLICATION NUMBER: US/09/390,721  
 ? EARLIER FILING DATE: 1999-09-07  
 ? EARLIER APPLICATION NUMBER: 60/100,134  
 ? EARLIER FILING DATE: 1998-09-14  
 ? NUMBER OF SEQ ID NOS: 6  
 ? SOFTWARE: PatentIn Ver. 2.0  
 ? SEQ ID NO 1  
 ? LENGTH: 5045  
 ? TYPE: DNA  
 ? ORGANISM: *Streptomyces avermitilis*  
 ? FEATURE:  
 ? NAME/KEY: CDS  
 ? LOCATION: (1112)..(2317)  
 ? OTHER INFORMATION: aver1 ORF  
 US-09-390-721-1  
 Alignment Scores:  
 Pred. No.: 3.47e+03 Length: 5045  
 Score: 36.00 Matches: 6  
 Percent Similarity: 80.00% Conservative: 2  
 Best Local Similarity: 60.00% Mismatches: 2  
 Query Match: 66.67% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-008-355-25 (1-10) x US-09-390-721-1 (1-5045)  
 QY 1 Thrglycylasnserylglyserprovalphe 10  
 Db 1479 ACTGGTGGCAACGCTGGTTACCCGCGCTGG 1508

RESULT 30  
US-09-390-721-3  
Sequence 3, Application US/09390721  
Patent No. 6197591  
GENERAL INFORMATION:  
APPLICANT: STUTZMAN-ENGWALL, KIM J.  
APPLICANT: PRICE, BRENDA S.  
TITLE OF INVENTION: STREPTOMYCES AVERMITILIS REGULATORY GENES FOR INCREASED  
FILE REFERENCE: PC9944A  
CURRENT APPLICATION NUMBER: US/09/390,721  
CURRENT FILING DATE: 1999-09-07  
EARLIER APPLICATION NUMBER: 60/100,134  
EARLIER FILING DATE: 1998-09-14  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 5045  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2314)..(3021)  
OTHER INFORMATION: aver2 ORF  
US-09-390-721-3

Alignment Scores:  
Pred. No.: 3,47e+03 Length: 5045  
Score: 36.00 Matches: 6  
Percent Similarity: 80.00% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 2  
Query Match: 66.67% Indels: 0  
DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-390-721-3 (1-5045)

QY 1 ThrglyglyAsnSerGlySerProValPhe 10  
|||||  
Db 1479 ACTGTGGCAACGCTGTTACCCGCCCTGG 1508

RESULT 31  
US-08-064-121-1  
Sequence 1, Application US/08064121  
Patent No. 5641664  
GENERAL INFORMATION:  
APPLICANT: D'HALLUIN, Kathleen  
APPLICANT: GOBEL, Elke  
TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/064,121  
FILING DATE: 24-MAY-1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 90403332.1  
FILING DATE: 23-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91401888.2  
FILING DATE: 08-JUL-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 010830-043  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5399 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: plasmid pDE108  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..451  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
FEATURE:  
NAME/KEY: -  
LOCATION: 452..1284  
OTHER INFORMATION: /label= 3553  
OTHER INFORMATION: /note= "3553 promoter sequence derived from  
OTHER INFORMATION: Cauliflower mosaic virus isolate cabbb-J1"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1285..2100  
OTHER INFORMATION: /label= NP11  
OTHER INFORMATION: /note= "coding sequence of neomycin  
OTHER INFORMATION: phosphotransferase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2101..3160  
OTHER INFORMATION: /label= 3'ocs  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from the  
OTHER INFORMATION: Agrobacterium T-DNA octopine synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 3161..5399  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
US-08-064-121-1

Alignment Scores:  
Pred. No.: 3,74e+03 Length: 5399  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x US-08-064-121-1 (1-5399)

QY 1 ThrglyglyAsnSerGlySerPro 8  
|||||  
Db 2712 ACTGAAGGAACTCCGCTTCCCGC 2735

RESULT 32  
US-08-478-015-1  
Sequence 1, Application US/08478015  
Patent No. 5712135  
GENERAL INFORMATION:  
APPLICANT: D'HALLUIN, Kathleen  
APPLICANT: GOBEL, Elke  
TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: P.O. Box 1404  
 City: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/478,015  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP 90403332.1  
 FILING DATE: 23-NOV-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP 91401888.2  
 FILING DATE: 08-JUL-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/064,121  
 FILING DATE: 23-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: McGowan, Malcolm K.  
 REGISTRATION NUMBER: 39,300  
 REFERENCE/DOCKET NUMBER: 010830-088  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 836-6620  
 TELEFAX: (703) 836-2021  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5399 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: circular  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: plasmid pDE108  
 FEATURE:  
 NAME/KEY: -  
 LOCATION: 1..451  
 OTHER INFORMATION: /label= pUC18  
 OTHER INFORMATION: /note= "pUC18 derived sequence"  
 FEATURE:  
 NAME/KEY: -  
 LOCATION: 452..1284  
 OTHER INFORMATION: /label= 3553  
 OTHER INFORMATION: /note= "3553 promoter sequence derived from Cauliflower mosaic virus isolate CabBB-J1"  
 FEATURE:  
 NAME/KEY: -  
 LOCATION: 1285..2100  
 OTHER INFORMATION: /label= NP11  
 OTHER INFORMATION: /note= "coding sequence of neomycine phosphotransferase gene"  
 FEATURE:  
 NAME/KEY: -  
 LOCATION: 2101..3160  
 OTHER INFORMATION: /label= 3'ocs  
 OTHER INFORMATION: /note= "3' regulatory sequence containing the polyadenylation site derived from the Agrobacterium T-DNA octopine synthase"  
 OTHER INFORMATION: gene"  
 FEATURE:  
 NAME/KEY: -  
 LOCATION: 3161..5399  
 OTHER INFORMATION: /label= pUC18  
 OTHER INFORMATION: /note= "pUC18 derived sequence"  
 US-08-478-015-1

Alignment Scores:  
 Pred. No.: 3,74e+03 Length: 5399  
 Score: 36.00 Matches: 7  
 Percent Similarity: 87.50% Conservative: 0  
 Best Local Similarity: 87.50% Mismatches: 1  
 Query Match: 66.67% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-008-355-25 (1-10) x US-08-478-015-1 (1-5399)  
 Qy 1 ThrclyGlyAsnSerGlySerPro 8  
 Db 2712 ACTGAGGAACTCGGTCCCG 2735  
 ||| |||||  
 RESULT 33  
 US-08-475-975-1  
 ; Sequence 1, Application US/08475975  
 ; Patent No. 6002070  
 ; GENERAL INFORMATION:  
 ; APPLICANT: D'HALLUIN, Kathleen  
 ; TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
 ; TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Burns, Doane, Swecker & Mathis  
 ; STREET: George Mason Bldg., Washington & Prince Sts.  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/475,975  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/064,121  
 ; FILING DATE: 24-MAY-1993  
 ; APPLICATION NUMBER: EP 90403332.1  
 ; FILING DATE: 23-NOV-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: EP 91401888.2  
 ; FILING DATE: 08-JUL-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Crane-Feury, Sharon E  
 ; REGISTRATION NUMBER: 36,113  
 ; REFERENCE/DOCKET NUMBER: 010830-043  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 836-6620  
 ; TELEFAX: (703) 836-2021  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5399 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; MOLECULE TYPE: DNA (genomic)  
 ; TOPOLOGY: circular  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: plasmid pDE108  
 ; FEATURE:  
 ; NAME/KEY: -  
 ; LOCATION: 1..451  
 ; OTHER INFORMATION: /label= pUC18  
 ; OTHER INFORMATION: /note= "pUC18 derived sequence"  
 ; NAME/KEY: -



LOCATION: 452..1284  
OTHER INFORMATION: /label= 3553  
OTHER INFORMATION: /note= "3553 promoter sequence derived from  
OTHER INFORMATION: Cauliflower mosaic virus isolate Cabbb-J1"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1285..2100  
OTHER INFORMATION: /label= NP11  
OTHER INFORMATION: /note= "coding sequence of neomycin  
OTHER INFORMATION: phosphotransferase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2101..3160  
OTHER INFORMATION: /label= 3'ocs  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from the  
OTHER INFORMATION: Agrobacterium T-DNA octopline synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 3161..5399  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
US-08-475-975-1  
Alignment Scores:  
Pred. No.: 3.74e+03 Length: 5399  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
DB: Gaps: 0  
US-10-008-355-25 (1-10) x US-08-475-975-1 (1-5399)  
Qy 1 ThrglyGlyAnserGlySerPro 8  
Db 2712 ACTGAAGGAACTCCGTTCCCG 2735  
RESULT 34  
US-09-084-889-1  
Sequence 1, Application US/09084889  
Patent No. 6074877  
GENERAL INFORMATION:  
APPLICANT: D'HALLOIN, Kathleen  
APPLICANT: GOBEL, Elke  
TITLE OF INVENTION: PROCESS FOR TRANSFORMING  
TITLE OF INVENTION: MONOCOTYLEDONOUS PLANTS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Burns, Doane, Swecker & Mathis  
STREET: George Mason Bldg., Washington & Prince Sts.  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/084,889  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/064,121  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 91401888.2  
FILING DATE: 08-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113

REFERENCE/DOCKET NUMBER: 010830-043  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5399 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: plasmid pDE108  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..451  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
FEATURE:  
NAME/KEY: -  
LOCATION: 452..1284  
OTHER INFORMATION: /label= 3553  
OTHER INFORMATION: /note= "3553 promoter sequence derived from  
OTHER INFORMATION: Cauliflower mosaic virus isolate Cabbb-J1"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1285..2100  
OTHER INFORMATION: /label= NP11  
OTHER INFORMATION: /note= "coding sequence of neomycin  
OTHER INFORMATION: phosphotransferase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2101..3160  
OTHER INFORMATION: /label= 3'ocs  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from the  
OTHER INFORMATION: Agrobacterium T-DNA octopline synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 3161..5399  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
US-09-084-889-1  
Alignment Scores:  
Pred. No.: 3.74e+03 Length: 5399  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
DB: Gaps: 0  
US-10-008-355-25 (1-10) x US-09-084-889-1 (1-5399)  
Qy 1 ThrglyGlyAnserGlySerPro 8  
Db 2712 ACTGAAGGAACTCCGTTCCCG 2735  
RESULT 35  
US-08-351-413-2  
Sequence 2, Application US/08351413  
Patent No. 5750867  
GENERAL INFORMATION:  
APPLICANT: Williams, Mark  
APPLICANT: Leemans, Jan  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESS: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church

STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/351,413  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/899,072  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/970,849  
FILING DATE: 03-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-102PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6555 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: plasmid pVE14 (replicable in E.coli)  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..396  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (397..751)  
OTHER INFORMATION: /label= 3'nos  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA nopaline synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (752..1024)  
OTHER INFORMATION: /label= barstar  
OTHER INFORMATION: /note= "coding region of the barstar gene of  
OTHER INFORMATION: Bacillus amyloliquefaciens"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (1025..1607)  
OTHER INFORMATION: /label= TA29  
OTHER INFORMATION: /note= "promoter derived from the TA29 gene of  
OTHER INFORMATION: Nicotiana tabacum"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1608..2440  
OTHER INFORMATION: /label= 35S3  
OTHER INFORMATION: /note= "35S3 promoter sequence derived from  
OTHER INFORMATION: cauliflower mosaic virus isolate cabdbb-j1"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2441..3256  
OTHER INFORMATION: /label= neo  
OTHER INFORMATION: /note= "coding region of the neomycine  
OTHER INFORMATION: phosphotransferase gene of Tn5"

FEATURE:  
NAME/KEY: -  
LOCATION: 3257..4315  
OTHER INFORMATION: /label= 3'ocs  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA octopine synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 4316..6555  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
US-08-351-413-2  
Alignment Scores:  
Pred. No.: 4.65e+03 Length: 6555  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Gaps: 0  
DB: 1  
US-10-008-355-25 (1-10) x US-08-351-413-2 (1-6555)  
QY 1 ThrGlyGlyAsnSerGlySerPro 8  
Db 3868 ACTGAGGAGACTCGGTTCCCG 3891  
RESULT 36  
US-09-025-583-2  
Sequence 2, Application US/09025583  
Patent No. 5977433  
GENERAL INFORMATION:  
APPLICANT: Williams, Mark  
APPLICANT: Leemans, Jan  
TITLE OF INVENTION: Maintenance of male-sterile plants  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road, Suite 500 East  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 2046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/025,583  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/351,413  
FILING DATE:  
APPLICATION NUMBER: US 07/899,072  
FILING DATE: 12-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/970,849  
FILING DATE: 03-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Svensson, Leonard R.  
REGISTRATION NUMBER: 30,330  
REFERENCE/DOCKET NUMBER: 2121-102PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6555 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: plasmid pVE14 (replicable in E.coli)  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..396  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (397..751)  
OTHER INFORMATION: /label= 3' nos  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA nopaline synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (752..1024)  
OTHER INFORMATION: /label= barstar  
OTHER INFORMATION: /note= "coding region of the barstar gene of  
OTHER INFORMATION: Bacillus amyloliquefaciens"  
FEATURE:  
NAME/KEY: -  
LOCATION: complement (1025..1607)  
OTHER INFORMATION: /label= TA29  
OTHER INFORMATION: /note= "promoter derived from the TA29 gene of  
OTHER INFORMATION: Nicotiana tabacum"  
FEATURE:  
NAME/KEY: -  
LOCATION: 1608..2440  
OTHER INFORMATION: /label= 35S3  
OTHER INFORMATION: /note= "35S3 promoter sequence derived from  
OTHER INFORMATION: cauliflower mosaic virus isolate CabD6-J1"  
FEATURE:  
NAME/KEY: -  
LOCATION: 2441..3256  
OTHER INFORMATION: /label= neo  
OTHER INFORMATION: /note= "coding region of the neomycin  
OTHER INFORMATION: /note= "coding region of the neomycin  
OTHER INFORMATION: phosphotransferase gene of Tn5"  
FEATURE:  
NAME/KEY: -  
LOCATION: 3257..4315  
OTHER INFORMATION: /label= 3' ocs  
OTHER INFORMATION: /note= "3' regulatory sequence containing the  
OTHER INFORMATION: polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA octopine synthase gene"  
FEATURE:  
NAME/KEY: -  
LOCATION: 4316..6555  
OTHER INFORMATION: /label= pUC18  
OTHER INFORMATION: /note= "pUC18 derived sequence"  
US-09-025-583-2  
Alignment Scores:  
Pred. No.: 4.65e+03 Length: 6555  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
Gaps: 0  
US-10-008-355-25 (1-10) x US-09-025-583-2 (1-6555)  
OY 1 ThrglyAsnSerglySerPro 8  
Db 3868 ACTGAGGCACTCGGTTCCCG 3891  
RESULT 37  
US-09-194-905-7

Sequence 7, Application US/0914905  
Patent No. 6306627  
GENERAL INFORMATION:  
APPLICANT: DECKER, Heinrich  
TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCSCENS  
TITLE OF INVENTION: GLA.O AND THEIR USE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/194,905  
FILING DATE: 29-JUL-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/EP97/02826  
FILING DATE: 30-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19622783.6  
FILING DATE: 07-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Granados, Patricia D.  
REGISTRATION NUMBER: 33,683  
REFERENCE/DOCKET NUMBER: 026083/0193  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6854 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-194-905-7  
Alignment Scores:  
Pred. No.: 4.88e+03 Length: 6854  
Score: 36.00 Matches: 6  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 85.71% Mismatches: 0  
Query Match: 66.67% Indels: 0  
Gaps: 0  
US-10-008-355-25 (1-10) x US-09-194-905-7 (1-6854)  
OY 2 glyglyAsnSerglySerPro 8  
Db 5946 GCGGCACTCGGTTCCCG 5966  
RESULT 38  
US-08-232-016-23  
Sequence 23, Application US/08232016  
Patent No. 5952547  
GENERAL INFORMATION:  
APPLICANT: CORNELIJSSEN, Marc  
APPLICANT: SOETAERT, Piet  
APPLICANT: STAM, Maiké  
APPLICANT: DOCKX, Jan  
APPLICANT: VAN AARSEN, Roel  
TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN  
TITLE OF INVENTION: PLANT CELLS  
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,016  
FILING DATE: 03-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 91402920.2  
FILING DATE: 30-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 92400820.4  
FILING DATE: 25-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K  
REGISTRATION NUMBER: P39,300  
REFERENCE/DOCKET NUMBER: 010830-049  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7566 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: plasmid DNA designated as pPS0212  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1785 /note= "Coding region of a  
OTHER INFORMATION: truncated modified bcl2 (cry1ab) gene, also designated as the  
OTHER INFORMATION: cry1ab6 gene."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1793..2026  
OTHER INFORMATION: /note= "3', regulatory sequence  
OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA gene 7."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2396..2921 /note= "35S promoter sequence  
OTHER INFORMATION: derived from Cauliflower mosaic virus."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2922..3581  
OTHER INFORMATION: /note= "Coding sequence of  
OTHER INFORMATION: chloramphenicol acetyl transferase gene."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 3582..4407  
OTHER INFORMATION: /note= "3', regulatory sequence  
OTHER INFORMATION: containing the polyadenylation site derived from Agrobacterium  
OTHER INFORMATION: T-DNA octopine synthase gene."  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 5600..6457  
OTHER INFORMATION: /note= "Sequence complementary to  
Patent No. 5952547  
OTHER INFORMATION: the coding sequence of the beta-lactamase gene."  
FEATURE:

```

1 NAME/KEY: misc_feature
2 LOCATION: 7071..7566
3 OTHER INFORMATION: /note="TR1" and TR2 promoter
4 OTHER INFORMATION: derived from Agrobacterium T-DNA (with modified leader with
5 OTHER INFORMATION: respect to sequence of pJDB84 of SEQ ID NO. 22."
6 US-08-232-016-23
7
8 Alignment Scores:
9
10 Pred. No.: 5.45e+03 Length: 7566
11 Score: 36.00 Matches: 7
12 Percent Similarity: 87.50% Conservative: 0
13 Best Local Similarity: 87.50% Mismatches: 1
14 Query Match: 66.67% Indels: 0
15 DB: 2 Gaps: 0
16
17 US-10-008-355-25 (1-10) x US-08-232-016-23 (1-7566)
18
19 QY 1 ThrGlyGlyAsnSerGlySerPro 8
20 ||| ||||| ||||| ||||| |||||
21 Db 1410 ACTGAAGGGAACCTCGGTTCCCGG 4163
22
23 RESULT 39
24 US-08-232-016-22
25 : Sequence 22, Application US/08232016
26 : Patent No. 5952547
27 :
28 : GENERAL INFORMATION:
29 : APPLICANT: CORNELISSEN, Marc
30 : APPLICANT: CORNETT, Piet
31 : APPLICANT: STAM, Maïke
32 : APPLICANT: DOCKX, Jan
33 : APPLICANT: VAN AARSEN, Roel
34 : TITLE OF INVENTION: MODIFIED GENES AND THEIR EXPRESSION IN
35 : TITLE OF INVENTION: PLANT CELLS
36 : NUMBER OF SEQUENCES: 23
37 : CORRESPONDENCE ADDRESS:
38 : ADDRESSEE: Burns, Doane, Swecker & Mathis
39 : STREET: P.O. Box 1404
40 : CITY: Alexandria
41 : STATE: Virginia
42 : COUNTRY: United States
43 : ZIP: 22313-1404
44 :
45 : COMPUTER READABLE FORM:
46 : MEDIUM TYPE: Floppy disk
47 : COMPUTER: IBM PC compatible
48 : OPERATING SYSTEM: PC-DOS/MS-DOS
49 : SOFTWARE: Patentin Release #1.0, Version #1.30
50 : CURRENT APPLICATION DATA:
51 : APPLICATION NUMBER: US/08/232.016
52 : FILING DATE: 03-NOV-1994
53 : CLASSIFICATION: 435
54 :
55 : PRIOR APPLICATION DATA:
56 : APPLICATION NUMBER: FR 91402920.2
57 : FILING DATE: 30-OCT-1991
58 : PRIOR APPLICATION DATA:
59 : APPLICATION NUMBER: GB 92400820.4
60 : FILING DATE: 25-MAR-1992
61 : ATTORNEY/AGENT INFORMATION:
62 : NAME: McGowan, Malcolm K
63 : REGISTRATION NUMBER: P39,300
64 : REFERENCE/DOCKET NUMBER: 010830-049
65 : TELECOMMUNICATION INFORMATION:
66 : TELEPHONE: (703) 836-6620
67 : TELEFAX: (703) 836-2021
68 : INFORMATION FOR SEQ ID NO. 22:
69 : SEQUENCE CHARACTERISTICS:
70 : LENGTH: 7639 base pairs
71 : TYPE: nucleic acid
72 : STRANDEDNESS: double
73 : TOPOLOGY: circular
74 : MOLECULE TYPE: DNA (genomic)
75 : ORGANISM: plasmid DNA designated as pJDB84
76 : FEATURE:

```

```

: NAME/KEY: CDS
: LOCATION: 1..1869
: OTHER INFORMATION: /note= "Coding region of a
: OTHER INFORMATION: truncated bt2 (cryIab) gene, also designated as the bt884 gen
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1877..2110
: OTHER INFORMATION: /note= "3' regulatory sequence
: OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteri
: OTHER INFORMATION: T-DNA gene 7."
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 2480..3005
: OTHER INFORMATION: /note= "35S promoter sequence
: OTHER INFORMATION: derived from Cauliflower mosaic virus."
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3006..3665
: OTHER INFORMATION: /note= "Coding sequence of
: OTHER INFORMATION: chloramphenicol acetyl transferase gene."
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 3666..4491
: OTHER INFORMATION: /note= "3' regulatory sequence
: OTHER INFORMATION: containing the polyadenylation site derived from Agrobacteri
: OTHER INFORMATION: T-DNA octopline synthase gene."
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 5684..6541
: OTHER INFORMATION: /note= "Sequence complementary to
: Patent No. 5952547
: OTHER INFORMATION: the coding sequence of the beta-lactamase gene."
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 7155..7639
: OTHER INFORMATION: /note= "TR1' and TR2' promoter
: OTHER INFORMATION: derived from Agrobacterium T-DNA."
: US-08-332-016-22
:
: Alignment Scores:
: Pred. No.: 5.51e+03 Length: 7639
: Score: 36.00 Matches: 7
: Percent Similarity: 87.50% Conservative: 0
: Best Local Similarity: 87.50% Mismatches: 1
: Query Match: 66.67% Indels: 0
: Gaps: 0
:
: US-10-008-355-25 (1-10) x US-08-232-016-22 (1-7639)
:
: QY 1 ThrGlyGlyAsnSerGlySerPro 8
: Db 4224 ACTGACGAGACTCGGTCCCG 4247
:
: RESULT 40
: US-08-676-169-1/c
: Sequence 1, Application US/08676169
: Patent No. 5773235
: GENERAL INFORMATION:
: APPLICANT: Chirnside, Ewan Douglas
: TITLE OF INVENTION: EQUINE ARTERITIS VIRUS PEPTIDES, ANTIBODIES
: TITLE OF INVENTION: AND THEIR USE IN A DIAGNOSTIC TEST
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NIXON & VANDERHAYE P.C.
: STREET: 1100 No. 5773235th Glebe Road, 8th Floor
: City: Arlington
: STATE: Virginia
: COUNTRY: U.S.A
: ZIP: 22201-4714
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/676.169
: FILING DATE: 31-Jul-96
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Crawford, Arthur R.
: REGISTRATION NUMBER: 25,327
: REFERENCE/DOCKET NUMBER: 1498-85
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 816-4006
: TELEFAX: (703) 816-4100
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12687 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: US-08-676-169-1
:
: Alignment Scores:
: Pred. No.: 9.66e+03 Length: 12687
: Score: 36.00 Matches: 7
: Percent Similarity: 87.50% Conservative: 0
: Best Local Similarity: 87.50% Mismatches: 1
: Query Match: 66.67% Indels: 0
: Gaps: 0
:
: US-10-008-355-25 (1-10) x US-08-676-169-1 (1-12687)
:
: QY 3 GlyAsnSerGlySerProValPhe 10
: Db 8775 GCGAATAGTGTCACCCCAATTTT 8752
:
: RESULT 41
: US-08-981-459-1/c
: Sequence 1, Application US/08981459
: Patent No. 6090390
: GENERAL INFORMATION:
: APPLICANT: CHIRNSIDE, Ewan Douglas
: TITLE OF INVENTION: Diagnostic Test For Equine Arteritis Virus
: TITLE OF INVENTION: Mediated Disease
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
: STREET: 400 Seventh Street, N.W.
: City: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/981.459
: FILING DATE: 19-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: HOLMAN, John C.
: REGISTRATION NUMBER: 22,769
: REFERENCE/DOCKET NUMBER: P61784US0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-638-6666
: TELEFAX: 202-393-5350
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12687 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

MOLECULE TYPE: cDNA  
US-08-981-459-1

Alignment Scores:  
Pred. No.: 9.66e+03 Length: 12687  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
Gaps: 0  
DB: 3

US-10-008-355-25 (1-10) x US-08-981-459-1 (1-12687)

Oy 3 GlyAsnSerGlySerProValPhe 10  
|||||

Db 8775 GCGAATAGTGTAGCCCAATTTT 8752

RESULT 42  
US-09-063-431A-1/c  
Sequence 1, Application US/09063431A  
Patent No. 6342222

GENERAL INFORMATION:  
APPLICANT: Chirnside, Ewan Douglas  
TITLE OF INVENTION: EQUINE ARTERITIS VIRUS PEPTIDES, ANTIBODIES  
TITLE OF INVENTION: AND THEIR USE IN A DIAGNOSTIC TEST  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NO. 6342222th Glebe Road, 8th Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A  
ZIP: 22201-4714

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/063,431A  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/676,169  
FILING DATE: 31-JUL-96  
ATTORNEY/AGENT INFORMATION:  
NAME: Crawford, Arthur R.  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 1498-85  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4006  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12687 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-09-063-431A-1

Alignment Scores:  
Pred. No.: 9.66e+03 Length: 12687  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
Gaps: 0

US-10-008-355-25 (1-10) x US-09-063-431A-1 (1-12687)

Oy 3 GlyAsnSerGlySerProValPhe 10  
|||||

Db 8775 GCGAATAGTGTAGCCCAATTTT 8752

RESULT 43  
US-08-673-768-1  
Sequence 1, Application US/08673768  
Patent No. 5952546

GENERAL INFORMATION:  
APPLICANT: Bedbrook, John R.  
APPLICANT: Dunsmuir, Pamela  
APPLICANT: Howie, William J.  
APPLICANT: Joe, Lawrence K.  
APPLICANT: Lee, Kathleen Y.  
TITLE OF INVENTION: Delayed Ripening Tomato Plants  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,768  
FILING DATE: 27-JUN-1996  
CLASSIFICATION: 800

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/000,721  
FILING DATE: 30-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 012176-005010US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15397 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-673-768-1

Alignment Scores:  
Pred. No.: 1.2e+04 Length: 15397  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
Gaps: 0

US-10-008-355-25 (1-10) x US-08-673-768-1 (1-15397)

Oy 1 ThrGlyGlyAsnSerGlySerPro 8  
|||||

Db 7549 ACTGAGGAGACTCCGTCCTCCCG 7572

RESULT 44  
US-08-673-768-1/c  
Sequence 1, Application US/08673768  
Patent No. 5952546

GENERAL INFORMATION:  
APPLICANT: Bedbrook, John R.  
APPLICANT: Dunsmuir, Pamela  
APPLICANT: Howie, William J.  
APPLICANT: Joe, Lawrence K.  
APPLICANT: Lee, Kathleen Y.

TITLE OF INVENTION: Delayed Ripening Tomato Plants  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,768  
FILING DATE: 27-JUN-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/000,721  
FILING DATE: 30-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 012176-0050100S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15397 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-673-768-1

Alignment Scores:  
Pred. No.: 1.2e+04 Length: 15397  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
Gaps: 0

US-10-008-355-25 (1-10) x US-08-673-768-1 (1-15397)

QY 1 ThrglyglaSnsrGlySerPro 8  
Db 13955 ACTGAGGGAAGCTCCGTTCCCGC 13932

RESULT 45  
5428147-1/c  
Patent No. 5428147  
APPLICANT: BARKER, RICHARD F.; KEMP, JOHN D.  
TITLE OF INVENTION: OCTOPINE T-DNA PROMOTERS  
NUMBER OF SEQUENCES: 17  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/91,538  
FILING DATE: 13-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 869,216  
FILING DATE: 13-APR-1992  
APPLICATION NUMBER: 869,216  
FILING DATE: 13-APR-1992  
APPLICATION NUMBER: 440,432  
FILING DATE: 21-NOV-1989  
APPLICATION NUMBER: 553,786  
FILING DATE: 19-NOV-1983  
APPLICATION NUMBER: 741,034  
FILING DATE: 06-AUG-1991  
APPLICATION NUMBER: 144,775  
FILING DATE: 20-JAN-1988

APPLICATION NUMBER: 485,614  
FILING DATE: 15-APR-1983  
APPLICATION NUMBER: 713,624  
FILING DATE: 10-JUN-1991  
APPLICATION NUMBER: 260,574  
FILING DATE: 21-OCT-1988  
APPLICATION NUMBER: 848,733  
FILING DATE: 01-APR-1986  
APPLICATION NUMBER: 535,354  
FILING DATE: 26-SEP-1983  
SEQ ID NO: 1:  
LENGTH: 24595  
5428147-1

Alignment Scores:  
Pred. No.: 2.01e+04 Length: 24595  
Score: 36.00 Matches: 7  
Percent Similarity: 87.50% Conservative: 0  
Best Local Similarity: 87.50% Mismatches: 1  
Query Match: 66.67% Indels: 0  
Gaps: 0

US-10-008-355-25 (1-10) x 5428147-1 (1-24595)

QY 1 ThrglyglaSnsrGlySerPro 8  
Db 12103 ACTGAGGGAAGCTCCGTTCCCGC 12080

RESULT 46  
US-09-091-814-104  
Sequence 104, Application US/09091814  
Patent No. 6218513  
GENERAL INFORMATION:  
APPLICANT: Anthony-Cahill, Spencer J.  
APPLICANT: Epp, Janet K.  
APPLICANT: Kerwin, Bruce A.  
APPLICANT: Ollins O., Peter  
APPLICANT: Mathews J., Anthony  
TITLE OF INVENTION: GLOBINS CONTAINING BINDING DOMAINS  
FILE REFERENCE: BYTB2005  
CURRENT APPLICATION NUMBER: US/09/091,814  
CURRENT FILING DATE: 1998-06-22  
NUMBER OF SEQ ID NOS: 116  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 104  
LENGTH: 45  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-091-814-104

Alignment Scores:  
Pred. No.: 27.1 Length: 45  
Score: 35.00 Matches: 6  
Percent Similarity: 88.89% Conservative: 2  
Best Local Similarity: 66.67% Mismatches: 1  
Query Match: 64.81% Indels: 0  
Gaps: 0

US-10-008-355-25 (1-10) x US-09-091-814-104 (1-45)

QY 1 ThrglyglaSnsrGlySerProVal 9  
Db 13 TCTGCTGCTCTGCTGCTTCCGTT 39

RESULT 47  
US-09-091-814-109/c  
Sequence 109, Application US/09091814  
Patent No. 6218513  
GENERAL INFORMATION:  
APPLICANT: Anthony-Cahill, Spencer J.

```

: APPLICANT: Epp, Janet K
: APPLICANT: Kerwin, Bruce A.
: APPLICANT: Olin O., Peter
: APPLICANT: Mathews J., Antony
: TITLE OF INVENTION: GLOBINS CONTAINING BINDING DOMAINS
: FILE REFERENCE: EXT2003
: CURRENT APPLICATION NUMBER: US/09/091,814
: CURRENT FILING DATE: 1998-06-22
: NUMBER OF SEQ ID NOS: 116
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 109
: LENGTH: 55
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence:
: OTHER INFORMATION: Oligonucleotide for P22 domain cassette
US-09-091-814-109

```

```

Alignment Scores:
Pred. No.: 34      Length: 55
Score: 35.00      Matches: 6
Percent Similarity: 88.89%      Conservative: 2
Best Local Similarity: 66.67%      Mismatches: 1
Query Match: 64.81%      Indels: 0
DB: 4      Gaps: 0

```

US-10-008-355-25 (1-10) x US-09-091-814-109 (1-55)

```

OY 1 ThrGlyGlyAsnSerGlySerProVal 9
Db 43 TCTGCGGCAATCGAGATCACCA 17

```

```

RESULT 48
US-09-364-539-58/C
: Sequence 58, Application US/09364539B
: Patent No. 6344321
: GENERAL INFORMATION:
: APPLICANT: Rabin, Ross
: APPLICANT: Lochrie, Michael
: APPLICANT: Janjic, Nebojsa
: APPLICANT: Gold, Larry
: TITLE OF INVENTION: Nucleic Acid Ligands Which Bind to Hepatocyte Growth
: TITLE OF INVENTION: Factor/Scatter Factor (HGF/SF) or Its Receptor C-Met
: FILE REFERENCE: NEX83
: CURRENT APPLICATION NUMBER: US/09/364,539B
: CURRENT FILING DATE: 1999-07-29
: NUMBER OF SEQ ID NOS: 192
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 58
: LENGTH: 62
: TYPE: RNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Synthetic
: OTHER INFORMATION: Sequence
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (1)..(61)
: OTHER INFORMATION: All pyrimidines are 2'F.
US-09-364-539-58

```

```

Alignment Scores:
Pred. No.: 38.8      Length: 62
Score: 35.00      Matches: 6
Percent Similarity: 87.50%      Conservative: 1
Best Local Similarity: 75.00%      Mismatches: 1
Query Match: 64.81%      Indels: 0
DB: 4      Gaps: 0

```

US-10-008-355-25 (1-10) x US-09-364-539-58 (1-62)

```

OY 1 ThrGlyGlyAsnSerGlySerPro 8

```

```

Db 50 TCTGCGGCAATCGAGATCACCA 27

```

```

RESULT 49
US-09-364-539-59/C
: Sequence 59, Application US/09364539B
: Patent No. 6344321
: GENERAL INFORMATION:
: APPLICANT: Rabin, Ross
: APPLICANT: Lochrie, Michael
: APPLICANT: Janjic, Nebojsa
: APPLICANT: Gold, Larry
: TITLE OF INVENTION: Nucleic Acid Ligands Which Bind to Hepatocyte Growth
: TITLE OF INVENTION: Factor/Scatter Factor (HGF/SF) or Its Receptor C-Met
: FILE REFERENCE: NEX83
: CURRENT APPLICATION NUMBER: US/09/364,539B
: CURRENT FILING DATE: 1998-07-29
: NUMBER OF SEQ ID NOS: 192
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 59
: LENGTH: 62
: TYPE: RNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Synthetic
: OTHER INFORMATION: Sequence
: NAME/KEY: modified_base
: LOCATION: (1)..(61)
: OTHER INFORMATION: All pyrimidines are 2'F.
US-09-364-539-59

```

```

Alignment Scores:
Pred. No.: 38.8      Length: 62
Score: 35.00      Matches: 6
Percent Similarity: 87.50%      Conservative: 1
Best Local Similarity: 75.00%      Mismatches: 1
Query Match: 64.81%      Indels: 0
DB: 4      Gaps: 0

```

US-10-008-355-25 (1-10) x US-09-364-539-59 (1-62)

```

OY 1 ThrGlyGlyAsnSerGlySerPro 8
Db 50 TCTGCGGCAATCGAGATCACCA 27

```

```

RESULT 50
US-09-342-681C-94/C
: Sequence 94, Application US/09342681C
: Patent No. 6355782
: GENERAL INFORMATION:
: APPLICANT: Zonana et al.
: TITLE OF INVENTION: Hypodiploid ectodermal dysplasia genes and proteins
: FILE REFERENCE: 52978
: CURRENT APPLICATION NUMBER: US/09/342,681C
: CURRENT FILING DATE: 1999-06-29
: PRIOR APPLICATION NUMBER: 60/092,279
: PRIOR FILING DATE: 1998-07-09
: PRIOR APPLICATION NUMBER: 60/112,366
: PRIOR FILING DATE: 1998-12-15
: NUMBER OF SEQ ID NOS: 123
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 94
: LENGTH: 425
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-342-681C-94

```

```

Alignment Scores:
Pred. No.: 333      Length: 425
Score: 35.00      Matches: 6
Percent Similarity: 88.89%      Conservative: 2
Best Local Similarity: 66.67%      Mismatches: 1

```



Query Match: 64.81% Indels: 0  
DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-342-681C-94 (1-425)

Oy 1 ThrGlyGlyAsnSerGlySerProVal 9  
|||||  
Db 335 ACAGGTGGCACCGTGGCACGCCCTCTC 309

Search completed: May 23, 2003, 14:23:04  
job time : 79 secs



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 12:11:33 ; Search time 1056 Seconds  
(Without alignments) 153.366 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TGCNCGSPVF 10

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xih  
-O=/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110501\_3097/app.query.fasta.1.199  
-DB=EST -OPMT=fastcap -SUFRTX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=1000  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355.@CGN.1.1456.@runat\_16052003\_110501\_3097 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV=TCPU=3  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELCP=6 -DELEXT=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estlun:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estlom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_lav:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vtc:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_man:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	47	87.0	318 9	AA411980
2	47	87.0	426 17	BH400391
3	47	87.0	608 10	BB622700
4	46	85.2	611 12	BF694219
5	45	83.3	371 17	AO699991
6	45	83.3	497 14	BQ327884
7	45	83.3	796 12	BF128636
8	45	83.3	914 12	BF128636
9	45	83.3	1108 12	BG115436
10	44	81.5	237 12	BF893687
11	44	81.5	279 10	AM060086
12	44	81.5	298 13	BJ099916
13	44	81.5	403 9	AA476126
14	44	81.5	420 14	W72307
15	44	81.5	432 10	AW743569
16	44	81.5	432 10	AW743601
17	44	81.5	447 9	AL117983
18	44	81.5	450 9	AA175055
19	44	81.5	491 9	AL034929
20	44	81.5	528 13	BI440854
21	44	81.5	713 12	BG751121
22	44	81.5	768 13	BG918747
23	44	81.5	812 17	AO912379
24	44	81.5	823 12	BF668889
25	44	81.5	857 13	BI695748
26	44	81.5	958 14	BQ881134
27	44	81.5	970 12	BE794754
28	44	81.5	3200 11	AK019499
29	43	79.6	278 10	BB176886
30	43	79.6	377 10	AM308476
31	43	79.6	392 10	AW315936
32	43	79.6	412 12	BG834820
33	43	79.6	422 13	BI830377
34	43	79.6	473 12	BG383001
35	43	79.6	483 13	BJ190590
36	43	79.6	584 12	BG834531
37	43	79.6	1466 13	BI545222
38	42	77.8	185 9	AI608160
39	42	77.8	192 13	BI783638
40	42	77.8	301 9	AA507304
41	42	77.8	319 9	AA596016
42	42	77.8	346 9	AA242434
43	42	77.8	352 9	AA447045
44	42	77.8	369 14	BM716334
45	42	77.8	375 12	BF083176
46	42	77.8	376 12	BF083167
47	42	77.8	376 12	BF083177
48	42	77.8	383 9	AA564287
49	42	77.8	389 9	AA069844
50	42	77.8	405 9	AI451721
51	42	77.8	426 9	AI661302
52	42	77.8	426 9	AA581647
53	42	77.8	456 9	AA664075
54	42	77.8	463 12	BG510048
55	42	77.8	473 12	BF397465
56	42	77.8	475 12	BF390358
57	42	77.8	484 12	BF467444
58	42	77.8	486 13	BM178660
59	42	77.8	495 12	AI029930
60	42	77.8	499 10	BF515765
61	42	77.8	514 10	BE330256
62	42	77.8	518 13	BM143769
63	42	77.8	565 10	AM666299
64	42	77.8	568 13	BM085217
				AA411980 zt65903.s
				BH400391 AC-ND-147
				BB622700 BB622700
				BF694219 602082842
				AO699991 HS-5334.A
				BQ327884 MR-R7004
				BF128636 602674594
				BF128636 601810895
				BG115436 602316560
				BF893687 QV1-M7016
				AM060086 687001G07
				BJ099916 BJ099916
				AA476126 vhl77d08.r
				W72307 z660908.s1
				AW743569 ur23d03.Y
				AW743601 ur23h03.Y
				AL117983 p9223a54
				AA175055 ms82901.r
				AL034929 m8708a52
				BI440854 lc77c12.Y
				BG751121 602728618
				BG918747 602819216
				AO912379 nbe000361
				BF668889 602123292
				BI695748 603347002
				BQ881134 AGENCOURT
				BE794754 601390584
				AK019499 Mus muscu
				BB176886 BI76886
				AM308476 3608 MARC
				AA315936 13723 MAR
				BG834820 353172 MA
				BI830377 603073426
				BG383001 300877 MA
				BJ190590 BJ190590
				BG834531 352788 MA
				BI545222 603187343
				AI608160 va14909.Y
				BI783638 kn35h01.Y
				AA507304 nh49d02.s
				AA596016 nbe6901.s
				AA242434 mx26c10.r
				AA447045 zw66a10.r
				BM716334 UT-E-CII-
				BF083176 QVO-CS001
				BF083167 QVO-CS001
				BF083177 QVO-CS001
				AA564287 nj05c11.s
				AA069844 zm15c10.r
				AI451721 mx26c10.x
				AI661302 va14909.x
				AA581647 nc85h10.s
				AA664075 UT-R-CO-g
				BG510048 sad27a11
				BF397465 UT-R-B52-
				BF390358 UT-R-CAI-
				BF467444 UT-M-CGP
				BM178660 sa153h01
				AI029930 UT-R-CO-j
				BF515765 UT-H-BW1-
				BE330256 sc75d09.Y
				BM143769 sa148e10
				AM666299 SK34h05.Y
				BM085217 sa133g11

C	65	42	77.8	576	17	BH300672	BH300672	CH230-200	138	41	75.9	565	17	TA14H020	AL452108 T. bruce1
	66	42	77.8	636	10	BE347015	BE347015	SP33H09.Y	139	41	75.9	572	17	AO160351	AO160351 mxxb0005B
	67	42	77.8	636	13	BM048712	BM048712	603625324	140	41	75.9	594	14	BO555979	BO555979 HA037D005-
	68	42	77.8	639	9	BI195610	BI195610	602754329	C 141	41	75.9	617	10	BE538609	BE538609 601065382
	69	42	77.8	640	9	AL537577	AL537577	AL537577	C 142	41	75.9	619	9	AL785210	AL785210 AL785210
	70	42	77.8	669	13	BI820440	BI820440	603034749	C 143	41	75.9	646	17	AG037756	AG037756 Pan trogl
	71	42	77.8	684	12	BE783786	BE783786	601471195	144	41	75.9	651	14	BQ397257	BQ397257 NISC_1926
	72	42	77.8	704	13	BI919278	BI919278	603183032	145	41	75.9	680	9	AL647551	AL647551 AL647551
	73	42	77.8	714	13	BB383120	BB383120	BB383120	146	41	75.9	680	10	AV751709	AV751709 AV751709
	74	42	77.8	724	13	BI756947	BI756947	603029154	147	41	75.9	682	17	AG083764	AG083764 Pan trogl
	75	42	77.8	735	13	BI909311	BI909311	603067276	148	41	75.9	690	14	BU038633	BU038633 DH02804 H
	76	42	77.8	752	12	BE743770	BE743770	601577877	149	41	75.9	692	9	AL654837	AL654837
	77	42	77.8	756	12	BE741863	BE741863	601593464	150	41	75.9	693	12	BF031211	BF031211 601560129
	78	42	77.8	760	13	BG387385	BG387385	602456150	151	41	75.9	704	9	AL647370	AL647370 AL647370
	79	42	77.8	760	13	BI769494	BI769494	603059039	152	41	75.9	764	10	AM174380	AM174380 AL62604.Y
	80	42	77.8	761	12	BF980935	BF980935	602304483	153	41	75.9	787	13	BI737468	BI737468 603358149
	81	42	77.8	774	13	BM049340	BM049340	603626296	154	41	75.9	801	9	AU002313	AU002313 AU002313
	82	42	77.8	781	9	AL530765	AL530765	AL530765	155	41	75.9	823	17	AZ701821	AZ701821 RPT1-23-2
	83	42	77.8	782	12	BG704790	BG704790	602688418	156	41	75.9	840	14	BO932460	BO932460 AGENCOURT
	84	42	77.8	787	13	BI829770	BI829770	603079841	157	41	75.9	871	14	BO716040	BO716040 AGENCOURT
	85	42	77.8	801	13	BG774896	BG774896	602649816	C 158	41	75.9	894	17	CNS03JFL	AL246790 Telradcon
	86	42	77.8	803	13	BI196351	BI196351	602755282	159	41	75.9	906	12	BG025138	BG025138 602276484
	87	42	77.8	804	12	BG746667	BG746667	602703976	C 160	41	75.9	916	14	BO619468	BO619468 RMOSE88A0
	88	42	77.8	817	12	BG702795	BG702795	602684682	161	41	75.9	960	14	BQ068715	BQ068715 AGENCOURT
	89	42	77.8	817	13	BI759253	BI759253	603042686	162	41	75.9	966	14	BQ435187	BQ435187 AGENCOURT
	90	42	77.8	822	13	BI758269	BI758269	603029920	163	41	75.9	966	14	BO644575	BO644575 AGENCOURT
	91	42	77.8	829	12	BG035239	BG035239	602324751	164	41	75.9	984	12	BF693852	BF693852 602082409
	92	42	77.8	837	12	BG716480	BG716480	602676205	165	41	75.9	1089	12	BE512670	BE512670 601171626
	93	42	77.8	838	13	BM017721	BM017721	603645090	166	41	75.9	1089	10	BE512670	BE512670 HMM021GA.
	94	42	77.8	845	12	BG742533	BG742533	602633385	167	41	75.9	1173	10	BE422171	BE422171 HMM021GA.
	95	42	77.8	851	9	AL521111	AL521111	AL521111	C 168	41	75.9	1271	13	BM563406	BM563406 AGENCOURT
	96	42	77.8	857	13	BI770448	BI770448	603055413	169	41	75.9	1481	13	BO960957	BO960957 AGENCOURT
	97	42	77.8	858	12	BG771219	BG771219	602719643	170	41	75.9	159	17	AZ031797	AZ031797 RPT1-23-3
	98	42	77.8	861	12	BF032529	BF032529	601453374	C 171	40	74.1	169	10	BB226987	BB226987 BB226987
	99	42	77.8	872	13	BI755987	BI755987	603030320	172	40	74.1	225	10	BB589647	BB589647 BB589647
	100	42	77.8	872	13	BM043661	BM043661	603620249	173	40	74.1	225	10	AA078678	AA078678 7701A03 C
	101	42	77.8	877	13	BI830579	BI830579	603073109	C 174	40	74.1	281	13	BI830035	BI830035 603080106
	102	42	77.8	882	12	BE959610	BE959610	601654576	C 175	40	74.1	297	10	BB718992	BB718992 BB718992
	103	42	77.8	894	12	BE868745	BE868745	601445981	176	40	74.1	313	17	AZ049915	AZ049915 GSGT1075
	104	42	77.8	916	12	BG823426	BG823426	602726884	C 177	40	74.1	319	10	BB534661	BB534661 BB534661
	105	42	77.8	926	12	BE792111	BE792111	601581670	C 178	40	74.1	325	9	AI026378	AI026378 TENU00589
	106	42	77.8	958	9	AL529745	AL529745	AL529745	C 179	40	74.1	349	10	AM325039	AM325039 TENU0433
	107	42	77.8	964	12	BG750659	BG750659	602708526	180	40	74.1	352	17	AQ907225	AQ907225 GSGT0927
	108	42	77.8	971	9	AL556984	AL556984	AL556984	C 181	40	74.1	383	9	AI065226	AI065226 TENU2110
	109	42	77.8	974	14	BM805164	BM805164	AGENCOURT	C 182	40	74.1	388	10	AM148054	AM148054 6030408.X
	110	42	77.8	976	13	BI825708	BI825708	603072693	183	40	74.1	391	9	AA413994	AA413994 vc64d04.s
	111	42	77.8	980	14	BM926742	BM926742	AGENCOURT	184	40	74.1	399	10	AM147471	AM147471 6030408.Y
	112	42	77.8	983	13	BM465277	BM465277	AGENCOURT	C 185	40	74.1	410	17	BM608562	BM608562 UP-357-7E
	113	42	77.8	995	14	BO958659	BO958659	AGENCOURT	C 186	40	74.1	418	13	BI187317	BI187317 a2004fs.T
	114	42	77.8	1028	14	BO921608	BO921608	AGENCOURT	187	40	74.1	420	9	AA831558	AA831558 oc64d04.s
	115	42	77.8	1039	14	BO063069	BO063069	AGENCOURT	C 188	40	74.1	424	12	BF231655	BF231655 d690a10.x
	116	42	77.8	1064	14	BO062263	BO062263	AGENCOURT	C 189	40	74.1	450	9	AA079786	AA079786 zmg3e01.r
	117	42	77.8	1065	13	BM563046	BM563046	AGENCOURT	C 190	40	74.1	451	17	BH192883	BH192883 TC3-66J9.
	118	42	77.8	1077	13	BM546794	BM546794	AGENCOURT	C 191	40	74.1	465	10	AM122726	AM122726 UI-M-BH2.
	119	42	77.8	1087	13	BM544010	BM544010	AGENCOURT	192	40	74.1	465	10	BE650601	BE650601 UI-M-BH2.
	120	42	77.8	1089	13	BM546202	BM546202	AGENCOURT	193	40	74.1	479	17	AZ776544	AZ776544 2M0010F13
	121	42	77.8	1127	13	BM545307	BM545307	AGENCOURT	C 194	40	74.1	480	10	AM122370	AM122370 UI-M-BH2.
	122	42	77.8	1139	13	BM450516	BM450516	AGENCOURT	C 195	40	74.1	480	10	AM401683	AM401683 UI-HF-BKO
	123	42	77.8	1140	14	BM924667	BM924667	AGENCOURT	C 196	40	74.1	482	9	AA475750	AA475750 vH23b12.r
	124	42	77.8	1151	13	BM543967	BM543967	AGENCOURT	C 197	40	74.1	482	12	BF024992	BF024992 dC82e01.x
	125	42	77.8	1647	14	BM903772	BM903772	AGENCOURT	C 198	40	74.1	483	12	BE960720	BE960720 UI-M-BG2-
	126	42	77.8	225	10	AW812682	AW812682	RCO-ST018	C 199	40	74.1	484	12	BG503240	BG503240 602551341
	127	41	75.9	320	14	W17876	W17876	mb73e03.r1	C 200	40	74.1	486	12	BG038423	BG038423 dG95b12.x
	128	41	75.9	453	9	AA161097	AA161097	z058b04.s	C 201	40	74.1	488	13	BJ052699	BJ052699 BJ052699
	129	41	75.9	467	10	AA161097	AA161097	Y197a08.s1	C 202	40	74.1	491	12	BF033084	BF033084 ux04021.Y
	130	41	75.9	467	10	BB702489	BB702489	BB702489	C 203	40	74.1	492	14	R36282	R36282 yG68c12.r1
	131	41	75.9	469	9	AL795494	AL795494	AL795494	C 204	40	74.1	494	9	AI037770	AI037770 ub51d09.r
	132	41	75.9	513	9	AA708882	AA708882	z194b03.s	C 205	40	74.1	494	12	BF659923	BF659923 mea25c01.
	133	41	75.9	541	13	BM116990	BM116990	L0844F08-	C 206	40	74.1	497	9	AI876857	AI876857 uJ35a09.Y
	134	41	75.9	543	17	TA300G0P	TA300G0P	T. bruce1	C 207	40	74.1	499	14	BO558705	BO558705 HA053c11.
	135	41	75.9	547	13	BM117404	BM117404	L0851B10-	C 208	40	74.1	501	13	BI055416	BI055416 BJ055416
	136	41	75.9	550	13	BM258134	BM258134	522831.MA	C 209	40	74.1	503	9	AI049350	AI049350 uc83f07.Y
	137	41	75.9	562	9	AT004740	AT004740	AT004740	210	40	74.1	506	9	AA764340	AA764340 vp02d05.r

211	40	74.1	506	17	BH829164	BH829164	BACPR28-F
C 212	40	74.1	509	12	BE957110	BE957110	UI-M-BG2
213	40	74.1	512	9	AA764333	AA764333	VP02C06..r
214	40	74.1	522	9	AA075141	AA075141	zm66f01.s
215	40	74.1	523	10	AM106781	AM106781	cm63f05.y
C 216	40	74.1	524	12	BE072322	BE072322	db56G03.x
C 217	40	74.1	529	10	BE366103	BE366103	601339208
C 218	40	74.1	539	9	A1951039	A1951039	wx62f08..x
C 219	40	74.1	539	13	BM026126	BM026126	fl74h08..y
C 220	40	74.1	551	12	BG554140	BG554140	de93b09..x
C 221	40	74.1	555	10	BE072022	BE072022	pm0-BR053
C 222	40	74.1	557	13	BI673971	BI673971	lE31901.Y
C 223	40	74.1	558	10	AM174529	AM174529	lJ05c08..Y
C 224	40	74.1	562	9	A1316985	A1316985	uJ25a10..Y
C 225	40	74.1	562	10	AM440755	AM440755	he05c06..x
C 226	40	74.1	570	9	AA548844	AA548844	nk28h12.s
C 227	40	74.1	573	17	BH819884	BH819884	BACPR12-I
C 228	40	74.1	577	13	BM195616	BM195616	C0313F10-
C 229	40	74.1	581	14	BQ667101	BQ667101	pB52d03.Y
C 230	40	74.1	585	13	BI450158	BI450158	dae75907..
C 231	40	74.1	588	10	AM106576	AM106576	um29h04..Y
C 232	40	74.1	593	14	BQ558706	BQ558706	H4053G11-
C 233	40	74.1	595	10	AV699648	AV699648	AV699648
C 234	40	74.1	598	13	BG915497	BG915497	602815721
C 235	40	74.1	601	12	BG534506	BG534506	de93b09..x
C 236	40	74.1	607	13	BJ050334	BJ050334	BJ050334
C 237	40	74.1	610	17	AQ0317850	AQ0317850	RPC111-96
C 238	40	74.1	612	10	BB637793	BB637793	BB637793
C 239	40	74.1	615	10	BB617618	BB617618	BB617618
C 240	40	74.1	623	17	AG100700	AG100700	pan tT091
C 241	40	74.1	624	10	AW761740	AW761740	uq23f02.Y
C 242	40	74.1	632	13	BJ056031	BJ056031	BJ056031
C 243	40	74.1	637	10	BB614486	BB614486	BB614486
C 244	40	74.1	639	13	BJ062466	BJ062466	BJ062466
C 245	40	74.1	643	9	AL705716	AL705716	DKEPR686F
C 246	40	74.1	657	12	BF843831	BF843831	MRI-HT118
C 247	40	74.1	658	13	BJ003023	BJ003023	BJ003023
C 248	40	74.1	668	10	BE372460	BE372460	601223905
C 249	40	74.1	673	17	AG034820	AG034820	pan tT091
C 250	40	74.1	685	9	AV084923	AV084923	AV084923
C 251	40	74.1	689	10	BE380796	BE380796	601270913
C 252	40	74.1	701	10	BE377865	BE377865	601230032
C 253	40	74.1	702	13	BJ089351	BJ089351	BJ089351
C 254	40	74.1	702	17	BH425720	BH425720	B0GBP76TR
C 255	40	74.1	703	13	BJ047396	BJ047396	BJ047396
C 256	40	74.1	704	9	AU177344	AU177344	AU177344
C 257	40	74.1	717	13	BI108424	BI108424	602902016
C 258	40	74.1	720	13	BJ057274	BJ057274	BJ057274
C 259	40	74.1	720	13	BJ083233	BJ083233	BJ083233
C 260	40	74.1	721	13	BJ087421	BJ087421	BJ087421
C 261	40	74.1	724	13	BJ056621	BJ056621	BJ056621
C 262	40	74.1	732	10	BE629092	BE629092	u311h08..Y
C 263	40	74.1	738	13	BG912352	BG912352	602808412
C 264	40	74.1	740	12	BE912627	BE912627	601663669
C 265	40	74.1	742	13	BG935778	BG935778	SSI-0083
C 266	40	74.1	751	10	BB643728	BB643728	BB643728
C 267	40	74.1	768	12	BF577452	BF577452	602092290
C 268	40	74.1	800	12	BF105161	BF105161	601822187
C 269	40	74.1	802	12	BE973357	BE973357	601652405
C 270	40	74.1	830	12	BE789906	BE789906	601476779
C 271	40	74.1	830	13	BG970599	BG970599	602837543
C 272	40	74.1	842	17	CNS023Y0	AL180105	Tetradon
C 273	40	74.1	851	12	BG110002	BG110002	602273933
C 274	40	74.1	851	17	BH718397	BH718397	BOHWZ83TR
C 275	40	74.1	864	12	BE958963	BE958963	601644690
C 276	40	74.1	867	12	BG754090	BG754090	602709626
C 277	40	74.1	877	12	BF701289	BF701289	602128377
C 278	40	74.1	881	14	BQ691318	BQ691318	AGENCOURT
C 279	40	74.1	903	17	AZ200073	AZ200073	SP_1040..B
C 280	40	74.1	906	17	CNS03A5A	AL234775	Tetradon
C 281	40	74.1	907	12	BF0675438	BQ675438	AGENCOURT
C 282	40	74.1	912	14	BF028443	BF028443	601763646
C 283	40	74.1	919	12	BF385766	BF385766	602045929
284	40	74.1	923	9	A1303445	A1303445	u160b01..x
C 285	40	74.1	932	17	CNS0385Y	AL232204	Tetradon
C 286	40	74.1	935	14	BQ946583	BQ946583	AGENCOURT
C 287	40	74.1	949	12	BE740409	BE740409	601595438
288	40	74.1	956	12	BG117067	BC117067	6023346237
C 289	40	74.1	956	14	BQ652033	BQ652033	AGENCOURT
C 290	40	74.1	959	12	BF785109	BF785109	602108287
C 291	40	74.1	977	17	CNS01SC0	AL165566	Tetradon
C 292	40	74.1	1003	17	CNS03PH0	AL125649	Tetradon
C 293	40	74.1	1013	17	CNS0258D	AL181750	Tetradon
C 294	40	74.1	1016	17	CNS02ACX	AL188394	Tetradon
C 295	40	74.1	1032	14	BQ652778	BQ652778	AGENCOURT
C 296	40	74.1	1042	14	BQ219078	BQ219078	AGENCOURT
C 297	40	74.1	1075	14	BQ065737	BQ065737	AGENCOURT
C 298	40	74.1	1096	13	BM547231	BM547231	AGENCOURT
C 299	40	74.1	1118	14	BQ893205	BQ893205	AGENCOURT
C 300	40	74.1	1118	14	BE372189	BE372189	601223508
C 301	40	74.1	1147	10	BE372189	BE372189	601223508
C 302	40	74.1	1525	12	BF685482	BF685482	602142424
C 303	40	74.1	1818	12	BE736063	BE736063	601305680
C 304	39	72.2	160	10	BB600303	BB600303	BB600303
C 305	39	72.2	161	17	AQ845990	AQ845990	LMAJFV1.1
C 306	39	72.2	206	10	BI157306	BI157306	BI157306
C 307	39	72.2	228	10	BB247782	BB247782	BB247782
C 308	39	72.2	235	10	BB381578	BB381578	BB381578
C 309	39	72.2	240	14	BU031760	BU031760	QU119h10..
C 310	39	72.2	256	17	CNS00XW2	AL095408	ArabiDops
C 311	39	72.2	261	17	BH225722	BH225722	1006128A0
C 312	39	72.2	263	17	AZ0604023	AZ0604023	RPC1-23-3
C 313	39	72.2	266	14	Z40650	A0500032	BG032..n
C 314	39	72.2	278	9	A1432312	A1432312	Lg54b10..x
C 315	39	72.2	282	12	BG231512	BE231512	na129C12..
C 316	39	72.2	285	10	BE363127	BE363127	DG1-9..C01
C 317	39	72.2	289	14	R09699	R09699	YF23c11..s1
C 318	39	72.2	299	9	A1551458	A1551458	mp87c04..Y
C 319	39	72.2	306	10	AM009848	AM009848	ws86c03..x
C 320	39	72.2	307	10	AM130265	AM130265	x148a01..x
C 321	39	72.2	309	9	AA582155	AA582155	ny36h09..s
C 322	39	72.2	309	14	BM713081	BM713081	UI-E-El0-
C 323	39	72.2	309	14	R55813	R55813	y992c08..s1
C 324	39	72.2	312	9	AU184607	AU184607	AU184607
C 325	39	72.2	322	9	A1202016	A1202016	q143a03..x
C 326	39	72.2	326	9	A1186537	A1186537	q435c11..x
C 327	39	72.2	326	10	BE221053	BE221053	hu20f08..x
C 328	39	72.2	329	14	R34017	R34017	yH83b08..r1
C 329	39	72.2	340	9	AA583703	AA583703	mb63a11.s
C 330	39	72.2	342	9	A1174537	A1174537	an45h09..s
C 331	39	72.2	344	13	BM485133	BM485133	539367..MA
C 332	39	72.2	348	14	N24540	N24540	yx75c05..s1
C 333	39	72.2	353	9	A1567783	A1567783	lr62e09..x
C 334	39	72.2	356	17	AZ283371	AZ283371	RPC1-23-1
C 335	39	72.2	359	9	A1129803	A1129803	q449e07..x
C 336	39	72.2	360	12	BG712742	BG712742	pg11n..PK0
C 337	39	72.2	362	9	AA599764	AA599764	ag35a04..s
C 338	39	72.2	363	9	AU032296	AU032296	AU032296
C 339	39	72.2	364	10	BB843299	BB843299	BB843299
C 340	39	72.2	370	17	FR0046953	AL443738	Fugu..rubr
C 341	39	72.2	374	9	A1277345	A1277345	gm48h12..x
C 342	39	72.2	380	12	BE857138	BE857138	7920B07..x
C 343	39	72.2	381	14	R00089	R00089	y673d08..s1
C 344	39	72.2	383	17	CNS00XOY	A1115186	qo94c10..x
C 345	39	72.2	386	9	AA351508	AA1095152	ArabiDops
C 346	39	72.2	386	9	BE919049	BE919049	CM0-NT013
C 347	39	72.2	388	12	BF606065	BF606065	273006..MA
C 348	39	72.2	391	12	BF432132	BF432132	na679f08..
C 349	39	72.2	395	12	BF432132	BF432132	na679f08..
C 350	39	72.2	404	14	T96720	T96720	ye51d04..r1
C 351	39	72.2	405	9	A1979180	A1979180	wr71f11..x
C 352	39	72.2	407	13	BM124214	BM124214	LO536h10-
C 353	39	72.2	407	14	N78814	N78814	z494G04..s1
C 354	39	72.2	410	14	H29018	H29018	ym59e04..s1
C 355	39	72.2	412	9	A1401634	A1401634	ch29e05..x
C 356	39	72.2	415	9	AA705042	AA705042	zj83c04..s

C 357	39	72.2	417	14	H41935	H41935	yo60h07.s1	C 430	39	72.2	525	9	AA588862	AA588862	no21h12.s
C 358	39	72.2	418	10	AW342022	AW342022	hb75c04.x	C 431	39	72.2	527	9	AA777525	AA777525	z194d01.s
C 359	39	72.2	419	10	AA708188	AA708188	zj76f08.s	C 432	39	72.2	528	12	BG899744	BG899744	HOA93-1-E
C 360	39	72.2	420	10	AW516964	AW516964	x05c10.x	C 433	39	72.2	528	17	BH225804	BH225804	1006128E0
C 361	39	72.2	420	14	N31566	N31566	yy19a11.s1	C 434	39	72.2	528	10	BE019677	BE019677	db28d11.x
C 362	39	72.2	421	10	AW383558	AW383558	xw70c06.x	C 435	39	72.2	535	14	BO143434	BO143434	fmb1c-pk0
C 363	39	72.2	421	10	AW489330	AW489330	ut-m-BH3-	C 436	39	72.2	535	17	BH225914	BH225914	1006129B0
C 364	39	72.2	424	10	AW489330	AW489330	ut-m-BH3-	C 437	39	72.2	537	9	A1754746	A1754746	ct29e12.x
C 365	39	72.2	425	10	BE575428	BE575428	dc55c03.x	C 438	39	72.2	539	12	BF192001	BF192001	243049.MA
C 366	39	72.2	425	14	R61308	R61308	yh07c07.s1	C 439	39	72.2	546	14	BO181960	BO181960	ut-H-EU0-
C 367	39	72.2	425	14	R61308	R61308	yh07c07.s1	C 440	39	72.2	549	10	BE377793	BE377793	601229949
C 368	39	72.2	427	13	BI892986	BI892986	saq81907.	C 441	39	72.2	549	12	BF705629	BF705629	229955.MA
C 369	39	72.2	430	10	A1554543	A1554543	tg22h07.x	C 442	39	72.2	551	17	BM992473	BM992473	ut-H-DP1-
C 370	39	72.2	430	10	BB784142	BB784142	bb784142	C 443	39	72.2	551	17	CNS02863	CNS02863	AL155932
C 371	39	72.2	433	13	A1002168	A1002168	og85a06.s	C 444	39	72.2	556	9	AA953814	AA953814	oo38d10.s
C 372	39	72.2	433	13	BI231229	BI231229	RE20464.5	C 445	39	72.2	560	17	BH819051	BH819051	BACP12-O
C 373	39	72.2	437	14	B0026711	B0026711	ut-1-BB0-	C 446	39	72.2	562	17	AQ708172	AQ708172	HS-5564-B
C 374	39	72.2	439	17	BH223705	BH223705	1006114E0	C 447	39	72.2	566	14	BO253385	BO253385	san69e02.
C 375	39	72.2	440	9	A1684334	A1684334	tc96a08.x	C 448	39	72.2	566	17	AQ798172	AQ798172	HS-3162.A
C 376	39	72.2	442	14	H99343	H99343	yx23e06.s1	C 449	39	72.2	568	17	AZ848497	AZ848497	2M0149G02
C 377	39	72.2	443	10	A1093328	A1093328	qb02d01.x	C 450	39	72.2	581	14	N30303	N30303	yw69h04.s1
C 378	39	72.2	444	10	AW583551	AW583551	1a01a09.y	C 451	39	72.2	581	17	BH547715	BH547715	BOHK647F
C 379	39	72.2	444	12	W79783	W79783	zd78e10.s1	C 452	39	72.2	584	12	BG212104	BG212104	RST31572
C 380	39	72.2	445	17	BH746464	BH746464	SALK 0427	C 453	39	72.2	586	17	AO162132	AO162132	mgxb0011H
C 381	39	72.2	445	14	N31823	N31823	yy18a09.s1	C 454	39	72.2	589	14	BM686920	BM686920	ut-E-CQ1-
C 382	39	72.2	447	14	AA503789	AA503789	nes1b10.s	C 455	39	72.2	589	17	BH383705	BH383705	AG-ND-158
C 383	39	72.2	448	14	R67877	R67877	yi28h09.r1	C 456	39	72.2	595	10	BE388768	BE388768	601283869
C 384	39	72.2	449	9	AA119882	AA119882	mp87c04.r1	C 457	39	72.2	599	10	BE064393	BE064393	RC4-BT031
C 385	39	72.2	451	9	A1669321	A1669321	wb85d01.x	C 458	39	72.2	602	17	B99067	B99067	CIT-HSP-228
C 386	39	72.2	452	9	AA911743	AA911743	o115g02.s	C 459	39	72.2	611	13	BJ067805	BJ067805	BJ067805
C 387	39	72.2	454	14	N20267	N20267	yx42c07.s1	C 460	39	72.2	615	17	BG086085	BG086085	H3121FP02-
C 388	39	72.2	457	14	AA219518	AA219518	zg99g11.r1	C 461	39	72.2	617	17	BH211048	BH211048	ATXOA44TR
C 389	39	72.2	461	10	A1305265	A1305265	qm08f02.x	C 462	39	72.2	625	14	BQ006952	BQ006952	ut-H-E11-
C 390	39	72.2	461	10	BE047127	BE047127	hg65a06.x	C 463	39	72.2	626	13	BJ0181935	BJ0181935	BO181935
C 391	39	72.2	462	17	BM988403	BM988403	ut-H-DH0-	C 464	39	72.2	628	17	BH823835	BH823835	BACP2-J1
C 392	39	72.2	463	14	BM988403	BM988403	ut-H-DH0-	C 465	39	72.2	629	14	BQ999819	BQ999819	OC623C07-
C 393	39	72.2	463	14	BM988403	BM988403	ut-H-DH0-	C 466	39	72.2	631	10	AV692428	AV692428	AV692428
C 394	39	72.2	465	9	A1624406	A1624406	ts29c11.s	C 467	39	72.2	632	10	BE359219	BE359219	DG1-39.C1
C 395	39	72.2	466	14	H17044	H17044	ym39g09.s1	C 468	39	72.2	633	13	BJ055105	BJ055105	BJ055105
C 396	39	72.2	468	9	A1827125	A1827125	md3c07.s	C 469	39	72.2	633	10	AM060106	AM060106	687002C08
C 397	39	72.2	471	12	BF251571	BF251571	EST418920	C 470	39	72.2	640	14	BF608230	BF608230	ppp2n-PK0
C 398	39	72.2	472	12	AM553116	AM553116	L0222B04-	C 471	39	72.2	640	12	W26293	W26293	25e4 Human
C 399	39	72.2	473	10	AM553116	AM553116	L0222B04-	C 472	39	72.2	642	13	BM538125	BM538125	ha90e08-9
C 400	39	72.2	474	14	BM153065	BM153065	TCBAP1Q13	C 473	39	72.2	642	17	BH835326	BH835326	BACP8-K0
C 401	39	72.2	474	14	BQ855285	BQ855285	QCB25L16.	C 474	39	72.2	644	17	AQ255380	AQ255380	mgxb0012H
C 402	39	72.2	474	14	BH223216	BH223216	1006111G1	C 475	39	72.2	644	17	BM1862335	BM1862335	603389114
C 403	39	72.2	476	13	BM439192	BM439192	IDLVR0235	C 476	39	72.2	645	13	BB611811	BB611811	BB611811
C 404	39	72.2	476	13	BM439192	BM439192	IDLVR0235	C 477	39	72.2	652	10	BJ037017	BJ037017	RJ037017
C 405	39	72.2	478	10	BB730227	BB730227	ut-R-CA1-	C 478	39	72.2	652	13	BH056246	BH056246	RPCT-24-3
C 406	39	72.2	478	12	BG376895	BG376895	ut-R-CA1-	C 479	39	72.2	654	17	BH873139	BH873139	hpa3802-B
C 407	39	72.2	482	17	BH822859	BH822859	BACPL8-A	C 480	39	72.2	672	17	BH873139	BH873139	hpa3802-B
C 408	39	72.2	487	14	H60714	H60714	yr53g11.s1	C 481	39	72.2	678	17	BQ772716	BQ772716	ut-H-FE0-
C 409	39	72.2	487	14	AO622512	AO622512	HS-2085_A	C 482	39	72.2	680	14	BQ772716	BQ772716	ut-H-FE0-
C 410	39	72.2	490	14	N45284	N45284	yz16b10.s1	C 483	39	72.2	681	12	BG287493	BG287493	602384503
C 411	39	72.2	496	13	BM402516	BM402516	SLA005B07	C 484	39	72.2	689	17	BH241025	BH241025	ATXOA44TR
C 412	39	72.2	497	17	AO359129	AO359129	HS-5029-B	C 485	39	72.2	689	17	AG063941	AG063941	pan.trog1
C 413	39	72.2	498	17	AO853343	AO853343	LMAUFL1-1	C 486	39	72.2	695	14	BM982240	BM982240	ut-CF-EN1
C 414	39	72.2	499	14	BO774452	BO774452	ut-H-E21-	C 487	39	72.2	695	14	BO021237	BO021237	ut-H-DH1-
C 415	39	72.2	502	12	BF657260	BF657260	OV2-22-F0	C 488	39	72.2	699	13	BM595014	BM595014	170006874
C 416	39	72.2	503	10	AW404639	AW404639	ut-HF-BL0	C 489	39	72.2	703	12	BH405291	BH405291	AG-ND-126
C 417	39	72.2	504	12	BF753574	BF753574	na655e01.	C 490	39	72.2	707	9	AL520544	AL520544	AL520544
C 418	39	72.2	505	10	BF432482	BF432482	na655e01.	C 491	39	72.2	721	14	BO990325	BO990325	AL520544
C 419	39	72.2	506	10	BE349057	BE349057	ht48e07.x	C 492	39	72.2	727	17	AG118426	AG118426	pan.trog1
C 420	39	72.2	508	17	BH216653	BH216653	1006044E1	C 493	39	72.2	730	17	CNS04V10	CNS04V10	Tetrarodon
C 421	39	72.2	509	17	AO131080	AO131080	HS-3032-B	C 494	39	72.2	731	17	AQ307107	AQ307107	HS-2028-A
C 422	39	72.2	509	17	BH223633	BH223633	1006114A0	C 495	39	72.2	750	12	BG722881	BG722881	602695368
C 423	39	72.2	512	9	AT923024	AT923024	wn81f09.x	C 496	39	72.2	751	14	BM982147	BM982147	ut-CF-EN1
C 424	39	72.2	513	9	AT097103	AT097103	qa03b05.x	C 497	39	72.2	754	14	BM980822	BM980822	ut-CF-EN1
C 425	39	72.2	514	12	BF656916	BF656916	wr70e03.x	C 498	39	72.2	759	12	BF868283	BF868283	963067C02
C 426	39	72.2	514	12	BF656916	BF656916	wr70e03.x	C 499	39	72.2	765	14	BM676794	BM676794	ut-E-E30-
C 427	39	72.2	515	14	BG474775	BG474775	z8b8h08.r1	C 500	39	72.2	779	12	BF628771	BF628771	HVSMED000
C 428	39	72.2	501	12	BG474775	BG474775	z8b8h08.r1	C 501	39	72.2	784	12	BF131228	BF131228	601819470
C 429	39	72.2	521	17	BH223743	BH223743	1006114G0	C 502	39	72.2	793	12	BG119939	BG119939	602352207

503	39	72.2	793	17	BH547803	BOGM20NR	576	39	72.2	1635	12	BF683082	BF683082	602139303
C 504	39	72.2	794	17	BH283232	CH230-39D	577	39	72.2	1691	13	BM546926	BM546926	AGENCOURT
C 505	39	72.2	809	10	BE641353	Cr12-3-A1	578	39	72.2	1727	12	BE962670	BE962670	60166042
C 506	39	72.2	809	12	BE911183	UI-H-ED1	579	39	72.2	1745	13	BM549664	BM549664	AGENCOURT
C 507	39	72.2	812	14	BO014129	UI-H-ED1	580	38-5	71.3	1007	12	BF235684	BF235684	602026823
C 508	39	72.2	831	10	BE567059	601341219	581	38	70.4	123	10	BE068470	BE068470	MK1-BT037
C 509	39	72.2	837	12	BE179508	601807383	582	38	70.4	136	10	AM326448	AM326448	19051 MAR
C 510	39	72.2	845	13	BE967120	601661193	583	38	70.4	185	13	BM258644	BM258644	523492 MA
C 511	39	72.2	852	13	BI834913	603090181	584	38	70.4	188	10	BM238355	BM238355	BB023835
C 512	39	72.2	856	12	BF212092	601813512	585	38	70.4	209	10	BE142155	BE142155	CM4-HT013
C 513	39	72.2	862	17	CNS01X56	AL171267	586	38	70.4	229	10	BR036376	BR036376	BM036376
C 514	39	72.2	866	13	BI644181	603207672	587	38	70.4	234	13	BS978983	BS978983	PM1-CN009
C 515	39	72.2	880	9	AL540202	AL540202	588	38	70.4	236	12	BF652885	BF652885	276593 MA
C 516	39	72.2	880	12	BG328101	602427153	589	38	70.4	239	9	AI142101	AI142101	OW61F04.x
C 517	39	72.2	883	12	BF107975	601824014	590	38	70.4	247	14	BM786799	BM786799	K-EST0065
C 518	39	72.2	883	14	BO651758	AGENCOURT	591	38	70.4	248	9	AA364088	AA364088	EST74633
C 519	39	72.2	886	12	BF974011	602243178	592	38	70.4	253	12	BF908129	BF908129	MR3-UT005
C 520	39	72.2	888	17	BH103815	RPC1-24-3	593	38	70.4	253	14	BO323632	BO323632	CM4-C1005
C 521	39	72.2	892	17	AO866902	nbebo0290	594	38	70.4	255	10	BE020931	BE020931	SM46F02.Y
C 522	39	72.2	896	12	BF162159	601770632	595	38	70.4	257	13	BJ510219	BJ510219	BJ510219
C 523	39	72.2	898	14	BO887628	AGENCOURT	596	38	70.4	274	12	BF440266	BF440266	BS2900020
C 524	39	72.2	899	17	CNS03AVT	AL235720	597	38	70.4	281	13	BM411534	BM411534	EST58580
C 525	39	72.2	905	12	BF106664	601823132	598	38	70.4	292	10	BR145685	BR145685	BM145685
C 526	39	72.2	909	12	BF307537	601893829	599	38	70.4	297	9	AA384732	AA384732	EST98284
C 527	39	72.2	920	14	BO672614	AGENCOURT	600	38	70.4	299	9	AV214101	AV214101	AV214101
C 528	39	72.2	922	12	BF304624	601888027	601	38	70.4	302	9	BE609906	BE609906	S946F09.Y
C 529	39	72.2	927	17	CNS072MT	AL426761	602	38	70.4	302	9	AA866025	AA866025	OH52602.s
C 530	39	72.2	929	12	BE896896	601439411	603	38	70.4	303	14	BO139736	BO139736	NR023G01P
C 531	39	72.2	930	12	BC9394839	602457202	604	38	70.4	305	14	H34308	H34308	EST111127.R
C 532	39	72.2	930	13	BM454610	AGENCOURT	605	38	70.4	309	10	AM485270	AM485270	64715 MAR
C 533	39	72.2	931	12	BF120418	601756272	606	38	70.4	312	10	BR404617	BR404617	BM404617
C 534	39	72.2	936	10	BE616172	601278736	607	38	70.4	313	10	AM804305	AM804305	OV0-UM009
C 535	39	72.2	937	12	BF304653	601888059	608	38	70.4	317	10	AV978172	AV978172	AV978172
C 536	39	72.2	940	11	AK010876	Mus muscu	609	38	70.4	330	12	BF149748	BF149748	uv77F04.Y
C 537	39	72.2	943	12	BE729744	601562937	610	38	70.4	331	10	BR530475	BR530475	BR530475
C 538	39	72.2	949	14	BO433428	AGENCOURT	611	38	70.4	342	13	BM116232	BM116232	L0833A02
C 539	39	72.2	950	17	CNS02098	AL173501	612	38	70.4	345	14	BO816268	BO816268	10030560
C 540	39	72.2	964	17	CNS03AVI	AL190050	613	38	70.4	349	12	BE994682	BE994682	UI-M-B21-
C 541	39	72.2	966	17	CNS02BMX	AL190050	614	38	70.4	351	14	RI1205	RI1205	YF41602.s1
C 542	39	72.2	972	12	BF308946	601889615	615	38	70.4	355	10	AM816101	AM816101	MR3-ST022
C 543	39	72.2	972	17	CNS0780H	AL134319	616	38	70.4	371	9	AA700230	AA700230	Z152F01.s
C 544	39	72.2	974	17	CNS078TD	T7 end of	617	38	70.4	372	14	T98839	T98839	Y662904.s1
C 545	39	72.2	975	14	BF183869	601842660	618	38	70.4	374	13	BI057437	BI057437	PM2-GN050
C 546	39	72.2	976	14	BO888822	AGENCOURT	619	38	70.4	375	10	AV667578	AV667578	AV667578
C 547	39	72.2	982	13	BI603266	603245122	620	38	70.4	389	12	BF601010	BF601010	265865 MA
C 548	39	72.2	984	12	BE909293	601502947	621	38	70.4	393	13	BI532963	BI532963	1024126B0
C 549	39	72.2	988	17	CNS04RLI	AL1304047	622	38	70.4	398	10	AV866944	AV866944	AV866944
C 550	39	72.2	993	13	BI457434	603185580	623	38	70.4	401	14	BU002779	BU002779	OGG32P11.
C 551	39	72.2	995	17	CNS03FNL	AL211914	624	38	70.4	402	10	AM816042	AM816042	MR3-ST022
C 552	39	72.2	1003	39	CNS06VOP	AL417407	625	38	70.4	411	13	BM286538	BM286538	526652 MA
C 553	39	72.2	1016	12	BE795432	601592989	626	38	70.4	414	12	BF442877	BF442877	260167 MA
C 554	39	72.2	1018	12	BF966784	602267877	627	38	70.4	414	14	CI9074	CI9074	CI9074
C 555	39	72.2	1020	12	BG113365	602283153	628	38	70.4	420	9	AU232945	AU232945	AD2332845
C 556	39	72.2	1026	13	BM423900	AGENCOURT	629	38	70.4	420	12	BF714792	BF714792	ma030312.
C 557	39	72.2	1046	14	BM803657	AGENCOURT	630	38	70.4	421	10	AM477778	AM477778	16644 MAR
C 558	39	72.2	1051	14	BO890728	AGENCOURT	631	38	70.4	422	10	AV593670	AV593670	AV593670
C 559	39	72.2	1052	17	CNS06H3N	AL1398505	632	38	70.4	422	10	BM849390	BM849390	BB849390
C 560	39	72.2	1054	14	BO930129	T3 end of	633	38	70.4	422	14	BM837436	BM837436	K-EST0113
C 561	39	72.2	1075	13	BO072772	AGENCOURT	634	38	70.4	425	9	AI148860	AI148860	EST247199
C 562	39	72.2	1100	13	BM555082	AGENCOURT	635	38	70.4	425	14	BO749965	BO749965	SNES74a74
C 563	39	72.2	1113	13	BM559970	AGENCOURT	636	38	70.4	427	10	AV406731	AV406731	AV406731
C 564	39	72.2	1118	12	BM490898	602520085	637	38	70.4	432	10	AM744395	AM744395	ur31907.Y
C 565	39	72.2	1132	13	BM560028	AGENCOURT	638	38	70.4	434	12	BF602278	BF602278	267471 MA
C 566	39	72.2	1147	13	BI834829	603090281	639	38	70.4	434	13	BI538399	BI538399	428965 MA
C 567	39	72.2	1155	14	BM926066	AGENCOURT	640	38	70.4	434	17	AO669964	AO669964	HS-5381.A
C 568	39	72.2	1172	12	BG475066	602491076	641	38	70.4	438	14	BO611683	BO611683	sap63F09.
C 569	39	72.2	1181	11	AK017532	Mus muscu	642	38	70.4	442	9	AI712041	AI712041	AI712041
C 570	39	72.2	1185	14	BO683255	AGENCOURT	643	38	70.4	442	12	BE953391	BE953391	UI-M-CD1-
C 571	39	72.2	1307	14	BO228859	AGENCOURT	644	38	70.4	447	17	AO511866	AO511866	HS 5123.A
C 572	39	72.2	1318	13	BM466762	AGENCOURT	645	38	70.4	454	10	AV596034	AV596034	AV596034
C 573	39	72.2	1459	13	BM019759	603648075	646	38	70.4	454	17	AZ060642	AZ060642	RPC1-23-4
C 574	39	72.2	1609	13	BM548156	AGENCOURT	647	38	70.4	455	10	BE208805	BE208805	bd15H07.x
C 575	39	72.2	1610	13	BI519240	603061982	648	38	70.4	455	13	BI946101	BI946101	su38g06.Y

c 649	38	70.4	457	10	AW986099	u f70h10.x	722	38	70.4	579	12	BG522312	BG522312 20-11 Ste
c 650	38	70.4	457	14	BO818306		723	38	70.4	580	10	AV609784	AV609784
c 651	38	70.4	457	14	BO9585	y f23c02.r1	724	38	70.4	580	14	BU005306	BU005306 OGG713.y
c 652	38	70.4	460	13	BIT25602		725	38	70.4	590	10	AM668416	CA_Ea001
c 653	38	70.4	464	12	BF923266	IL2-NT09	726	38	70.4	590	12	BG646123	EST507742
c 654	38	70.4	467	12	BF248628	NCEST3a01	727	38	70.4	594	17	BH008975	BH008975 e f22c08.x
c 655	38	70.4	468	14	BO758402	EBma05_SQ	728	38	70.4	598	10	AV833627	AV833627
c 656	38	70.4	468	14	BO998436	BO998436	729	38	70.4	599	10	BE367949	601221819
c 657	38	70.4	469	9	AL514893	AL514893	730	38	70.4	599	14	BO860125	BO860125 OGC14015
c 658	38	70.4	470	9	AM461830	AM461830 BP230006B	731	38	70.4	600	10	BB229663	BB229663
c 659	38	70.4	474	9	AA456472	AA564472 z x74g12.f	732	38	70.4	601	13	BIT83733	BIT83733 603086406
c 660	38	70.4	474	13	BM255676	BM255676 517853.MA	733	38	70.4	601	14	BO780741	BO780741 UT-R-F0-
c 661	38	70.4	475	10	AM243413	AM243413 xm95a10.x	734	38	70.4	602	9	AI1795612	AI1795612 614004C04
c 662	38	70.4	475	14	BO872635	BO872635 OGI16D17.	735	38	70.4	602	14	BO6337126	BO6337126 he06a09.y
c 663	38	70.4	476	10	AM367910	AM367910 MR0-HT016	736	38	70.4	604	17	AZ469920	AZ469920 im028a0P7
c 664	38	70.4	477	14	BO995549	BO995549 OGG10F02.	737	38	70.4	607	14	BM724394	UT-E-E01-
c 665	38	70.4	481	17	BH858607	BH858607 S3_008b.t	738	38	70.4	607	17	AG156976	AG156976 Pan t r091
c 666	38	70.4	482	12	BF604341	BF604341 270134.MA	739	38	70.4	608	14	BO996564	BO996564 OGC13D01.
c 667	38	70.4	483	14	BM987967	BM987967 UI-H-C00-	740	38	70.4	608	14	BO996536	BO996536 OGC14D14.
c 668	38	70.4	487	12	BF070843	BF070843 s126d06.y	741	38	70.4	610	13	BI520458	BI520458 603071648
c 669	38	70.4	487	14	BO638301	BO638301 hd20h06.y	742	38	70.4	612	14	BO863400	BO863400 OGC23M13.
c 670	38	70.4	487	17	AO391975	AO391975 CITRf-EI-	743	38	70.4	616	17	AZ700721	AZ700721 RPCI-23-2
c 671	38	70.4	489	13	BM093548	BM093548 sa110d07.	744	38	70.4	618	14	BO858649	BO858649 OGC10N12.
c 672	38	70.4	490	9	AI180079	AI180079 zp13104.f	745	38	70.4	619	13	BM316800	f w77f02.x
c 673	38	70.4	490	12	BF604345	BF604345 270139.MA	746	38	70.4	623	13	BO874915	BO874915
c 674	38	70.4	490	14	N22432	N22432 yw39f02.s1	747	38	70.4	624	12	BF694798	BF694798 602080683
c 675	38	70.4	491	14	BO118763	BO118763 EST604339	748	38	70.4	624	13	BC913270	BC913270 602811956
c 676	38	70.4	492	9	AI085588	AI085588 oy88d10.x	749	38	70.4	625	13	BM417093	BM417093 952003E04
c 677	38	70.4	494	10	AV735676	AV735676 AV735676	750	38	70.4	625	13	BO996336	BO996336 OGC12119.
c 678	38	70.4	495	12	BF603287	BF603287 268774.MA	751	38	70.4	628	13	BIT62502	BIT62502 603380625
c 679	38	70.4	495	12	BM419155	BM419155 R010A09.O	752	38	70.4	629	9	AI1967199	AI1967199 614050D06
c 680	38	70.4	497	9	AI1562442	AI1562442 TENS1976	753	38	70.4	634	14	BO862397	BO862397 QGC20P19.
c 681	38	70.4	499	14	BO996953	BO996953 OGC14E09.	754	38	70.4	634	14	W67725	w67725 zd42h04.r1
c 682	38	70.4	501	10	AM774454	AM774454 EST333605	755	38	70.4	635	14	BO999129	BO999129 OGC21D08.
c 683	38	70.4	502	9	AI038262	AI038262 oy85f08.x	756	38	70.4	635	14	BU002456	BU002456 OGC13G14.
c 684	38	70.4	505	13	BM090019	BM090019 503908.MA	757	38	70.4	637	14	BO997017	BO997017 OGC14H18.
c 685	38	70.4	510	17	BH051261	BH051261 RPCI-24-3	758	38	70.4	640	14	BO875905	BO875905 OGC19H08.y
c 686	38	70.4	514	12	BF852395	BF852395 MR3-EN008	759	38	70.4	640	17	AZ448049	AZ448049 IM0245F04
c 687	38	70.4	515	10	BB757940	BB757940 BR757940	760	38	70.4	640	17	BH488336	BH488336 BOGHU21TF
c 688	38	70.4	519	12	BF635343	BF635343 NF060E01D	761	38	70.4	641	17	AZ272846	RPCI-23-1
c 689	38	70.4	520	14	BM866775	BM866775 MFG0606K1	762	38	70.4	647	14	BO864722	BO864722 OGC27H19.
c 690	38	70.4	521	14	BM837701	BM837701 UI-M-CD1-	763	38	70.4	649	12	BE959296	BE959296 601654204
c 691	38	70.4	525	13	BIT75716	BIT75716 468400.MA	764	38	70.4	651	10	AM586592	AM586592 EST318152
c 692	38	70.4	526	12	BF853451	BF853451 MR3-EN008	765	38	70.4	653	14	BO198891	BO198891 UI-R-DO1-
c 693	38	70.4	527	12	BF080519	BF080519 231444.MA	766	38	70.4	659	13	BC961644	BC961644 602826643
c 694	38	70.4	527	12	BF388800	BF388800 UI-R-B52-	767	38	70.4	659	13	BI161479	BI161479 602865468
c 695	38	70.4	528	13	BJ067114	BJ067114 BJ067114	768	38	70.4	659	17	BH313283	BH313283 CH230-101
c 696	38	70.4	530	13	BM272378	BM272378 1940h02.y	769	38	70.4	661	12	BE970020	BE970020 601680130
c 697	38	70.4	531	13	BM665570	BM665570 UI-E-CL1-	770	38	70.4	663	17	AG060448	AG060448 Pan t r091
c 698	38	70.4	532	10	AV599547	AV599547	771	38	70.4	664	11	AV110476	z ea may5
c 699	38	70.4	535	12	BE757678	BE757678 212188.MA	772	38	70.4	666	12	BG854638	BG854638 102404080
c 700	38	70.4	535	17	AZ918312	AZ918312 100600A0	773	38	70.4	667	12	BG474420	BG474420 602517141
c 701	38	70.4	536	14	BM711755	BM711755 UI-E-CL1-	774	38	70.4	667	13	BI144599	BI144599 602909475
c 702	38	70.4	537	12	BF072877	BF072877 219373.MA	775	38	70.4	667	14	BO996744	BO996744 OGC13L07.
c 703	38	70.4	539	17	AZ356324	AZ356324 IM0097C23	776	38	70.4	668	12	BG851865	BG851865 10240332E1
c 704	38	70.4	540	9	AA042442	AA042442 24747.CD4	777	38	70.4	669	10	AV544228	AV544228 AV544228
c 705	38	70.4	544	13	BM280531	BM280531 K102a04.y	778	38	70.4	670	10	BB625930	BB625930
c 706	38	70.4	545	13	BI681216	BI681216 460544.MA	779	38	70.4	670	14	BO998276	BO998276 OGC19A21.
c 707	38	70.4	547	17	AZ408042	AZ408042 IM0179M01	780	38	70.4	671	10	BU004821	BU004821 OGC6F04.y
c 708	38	70.4	551	17	AZ812271	AZ812271 2M0078P15	781	38	70.4	673	12	BF452220	BF452220 u z86a07.y
c 709	38	70.4	552	17	AO493780	AO493780 HS_5071_A	782	38	70.4	676	13	BI131112	BI131112 EST531286
c 710	38	70.4	553	9	AA187012	AA187012 zp72h06.f	783	38	70.4	677	10	BE383462	BE383462 Pan t r091
c 711	38	70.4	557	17	AZ416633	AZ416633 IM0192A07	784	38	70.4	682	17	AG143750	AG143750
c 712	38	70.4	561	14	BO783635	BO783635 fab32a08.	785	38	70.4	684	10	AM556607	AM556607 L0270G08-
c 713	38	70.4	562	17	AZ614060	AZ614060 IM0442007	786	38	70.4	685	14	BU005662	BU005662 OGC8B04.y
c 714	38	70.4	565	17	AZ386667	AZ386667 IM0145F18	787	38	70.4	687	14	BO863976	BO863976 OGC25G07.
c 715	38	70.4	565	13	BM255618	BM255618 517405.MA	788	38	70.4	690	14	BU011259	OCG15L16.
c 716	38	70.4	572	10	AM690229	AM690229 NF030F04S	789	38	70.4	692	12	BG432331	BG432331 602496650
c 717	38	70.4	572	14	BU002946	OGG3JJ24.	790	38	70.4	692	14	BO862351	OGC20N20.
c 718	38	70.4	572	17	FR0013931	AL005175 F.rubrtipe	791	38	70.4	694	12	BG394689	BG394689 602456644
c 719	38	70.4	574	17	AZ456506	AZ456506 IM0259F17	792	38	70.4	696	17	AG109641	Pan t r091
c 720	38	70.4	575	12	BG362879	BG362879 sac13a09.	793	38	70.4	697	10	BB650111	BB650111
c 721	38	70.4	578	17	AO526866	HS_5219_A	794	38	70.4	700	12	BG856843	BG856843 1024048D1



795	38	70.4	700	14	B0871538	B0871538	OCG12C04.	868	38	70.4	908	13	B1859050	B1859050	603387995
796	38	70.4	702	14	B0996109	B0996109	OCG11014.	869	38	70.4	912	10	AM053304	AM053304	L30-1500T
797	38	70.4	702	14	B0998570	B0998570	OCG19022.	870	38	70.4	917	12	BF301395	BF301395	602029850
798	38	70.4	704	13	B1924777	B1924777	EST544666	871	38	70.4	918	14	B0212353	B0212353	AGENCOURT
799	38	70.4	705	14	B0859127	B0859127	OCG12C16.	872	38	70.4	918	14	B0277101	B0277101	AGENCOURT
800	38	70.4	707	10	AM184500	AM184500	LT15609.-Y	873	38	70.4	918	14	B0680661	B0680661	AGENCOURT
801	38	70.4	707	14	B0995774	B0995774	OCG10023.	874	38	70.4	919	12	BF103435	BF103435	601646877
802	38	70.4	709	14	B0512451	B0512451	AG086430 Pan Trogl	875	38	70.4	920	12	BF795884	BF795884	602259126
803	38	70.4	711	17	AG086430	AG086430	Pan Trogl	876	38	70.4	922	12	BF967386	BF967386	602287282
804	38	70.4	714	14	B0578591	B0578591	WHE0307.B	877	38	70.4	922	12	BG286144	BG286144	602383354
805	38	70.4	716	10	BE275519	BE275519	601121368	878	38	70.4	924	10	BE542478	BE542478	601053852
806	38	70.4	719	13	B1696373	B1696373	603345585	879	38	70.4	929	12	BG390666	BG390666	602415289
807	38	70.4	719	17	AZ109193	AZ109193	RPCI-23-4	880	38	70.4	931	14	B0711219	B0711219	AGENCOURT
808	38	70.4	720	14	B0865571	B0865571	OCG5F08.-Y	881	38	70.4	933	12	BF027287	BF027287	601671608
809	38	70.4	726	14	B0865568	B0865568	OCG24D20.	882	38	70.4	933	14	B0943158	B0943158	AGENCOURT
810	38	70.4	727	14	B0873015	B0873015	OCG11F02.	883	38	70.4	936	17	CNS04DKT	AL256598	Tetradon
811	38	70.4	728	14	B0865590	B0865590	OCG5F17.-Y	884	38	70.4	937	12	B0226598	B0226598	AGENCOURT
812	38	70.4	729	14	B0862331	B0862331	OCG20M24.	885	38	70.4	939	12	BG404134	BG404134	602420045
813	38	70.4	729	17	AZ858365	AZ858365	2M0163E05	886	38	70.4	942	14	B0676220	B0676220	AGENCOURT
814	38	70.4	730	14	B0865540	B0865540	OCG5D12.-Y	887	38	70.4	943	14	B0895960	B0895960	AGENCOURT
815	38	70.4	732	14	B0860940	B0860940	OCG17C01.	888	38	70.4	946	12	BG120258	BG120258	602353645
816	38	70.4	735	14	B0005119	B0005119	OCG7C24.-Y	889	38	70.4	949	14	B0958532	B0958532	AGENCOURT
817	38	70.4	737	14	B0004442	B0004442	OCG5D11.-Y	890	38	70.4	952	12	BG339950	BG339950	602438085
818	38	70.4	738	14	B0860102	B0860102	OCG14N15.	891	38	70.4	953	17	AG162989	AG162989	Pan Trogl
819	38	70.4	739	14	B0997315	B0997315	OCG16B22.	892	38	70.4	957	14	B0942610	B0942610	AGENCOURT
820	38	70.4	739	14	B0012236	B0012236	OCJ1G05.-Y	893	38	70.4	961	14	B0954384	B0954384	AGENCOURT
821	38	70.4	746	12	BE785904	BE785904	601477904	894	38	70.4	964	17	CNS03P7B	AL254288	Tetradon
822	38	70.4	750	17	BH503412	BH503412	BG06M34TF	895	38	70.4	967	17	AZ686985	AZ686985	ENTJ13TR
823	38	70.4	756	13	B1935558	B1935558	EST555447	896	38	70.4	970	13	BM556617	BM556617	AGENCOURT
824	38	70.4	757	9	AL525313	AL525313	PL1-1-3.H	897	38	70.4	970	14	B0882666	B0882666	AGENCOURT
825	38	70.4	757	17	AQ0869225	AQ0869225	nbe003C40	898	38	70.4	971	12	BG844642	BG844642	1024007C0
826	38	70.4	758	10	BB023186	BB023186	B044217	899	38	70.4	975	12	BE795949	BE795949	601591112
827	38	70.4	760	14	B0434217	B0434217	AGENCOURT	900	38	70.4	976	12	BG025065	BG025065	602276408
828	38	70.4	774	12	BE272078	BE272078	GA_EB001	901	38	70.4	976	13	B1082383	B1082383	602877582
829	38	70.4	776	12	BE294246	BE294246	001PBH07	902	38	70.4	977	14	BQ422706	BQ422706	AGENCOURT
830	38	70.4	777	10	BE613779	BE613779	601504584	903	38	70.4	979	17	CNS05HNP	AL37822	Tetradon
831	38	70.4	778	13	B1970794	B1970794	GM830011B	904	38	70.4	986	14	BQ944124	BQ944124	AGENCOURT
832	38	70.4	780	17	AZ192509	AZ192509	SP_1021_B	905	38	70.4	990	12	BG675264	BG675264	602621553
833	38	70.4	783	12	BG831245	BG831245	602676647	906	38	70.4	993	12	BG293279	BG293279	602390857
834	38	70.4	786	12	BE125962	BE125962	601762985	907	38	70.4	997	13	BI407441	BI407441	602919229
835	38	70.4	790	12	BE778680	BE778680	601463817	908	38	70.4	997	14	BQ673176	BQ673176	AGENCOURT
836	38	70.4	795	12	BG393116	BG393116	602411385	909	38	70.4	999	12	BG667071	BG667071	DRABVG04
837	38	70.4	796	17	BH654179	BH654179	BOMBL48TR	910	38	70.4	1003	14	B0885072	B0885072	AGENCOURT
838	38	70.4	798	12	BG402701	BG402701	602418238	911	38	70.4	1004	17	CNS03COM	AL238073	Tetradon
839	38	70.4	803	12	BG036609	BG036609	602326803	912	38	70.4	1005	13	BM029456	BM029456	IPSK00128
840	38	70.4	809	13	BI100837	BI100837	602886146	913	38	70.4	1010	14	BQ926265	BQ926265	AGENCOURT
841	38	70.4	809	13	BM228819	BM228819	KX268E12-	914	38	70.4	1010	17	CNS03LTM	AL249125	Tetradon
842	38	70.4	815	14	BQ641976	BQ641976	AGENCOURT	915	38	70.4	1012	13	BM542175	BM542175	AGENCOURT
843	38	70.4	830	12	BG401091	BG401091	602465157	916	38	70.4	1016	17	CNS04Q50	AL302157	Tetradon
844	38	70.4	831	10	BE277174	BE277174	601178520	917	38	70.4	1023	9	AL551258	AL551258	AGENCOURT
845	38	70.4	836	14	BQ212500	BQ212500	AGENCOURT	918	38	70.4	1029	13	BM476107	BM476107	AGENCOURT
846	38	70.4	841	17	AO745661	AO745661	HS_2273_A	919	38	70.4	1030	17	CNS028LV	AL166124	Tetradon
847	38	70.4	848	12	BG474227	BG474227	602516846	920	38	70.4	1050	12	BG254425	BG254425	602354972
848	38	70.4	850	12	BG756284	BG756284	602713612	921	38	70.4	1060	17	CNS04UKI	AL307882	Tetradon
849	38	70.4	852	13	BI107285	BI107285	602894284	922	38	70.4	1062	14	BM921475	BM921475	AGENCOURT
850	38	70.4	857	12	BG210476	BG210476	RST30101	923	38	70.4	1077	14	BQ683924	BQ683924	AGENCOURT
851	38	70.4	858	12	BF346258	BF346258	6020818447	924	38	70.4	1079	13	BI197251	BI197251	602757995
852	38	70.4	859	13	BG914628	BG914628	602813387	925	38	70.4	1096	14	BM910951	BM910951	AGENCOURT
853	38	70.4	866	9	AU124976	AU124976	AU124976	926	38	70.4	1109	13	BI259074	BI259074	602971769
854	38	70.4	867	12	BG075614	BG075614	H3149D04-	927	38	70.4	1115	13	BM480299	BM480299	AGENCOURT
855	38	70.4	869	17	CNS04SCY	AL305035	Tetradon	928	38	70.4	1119	13	BM541859	BM541859	AGENCOURT
856	38	70.4	874	14	BQ691878	BQ691878	AGENCOURT	929	38	70.4	1127	14	BM804216	BM804216	AGENCOURT
857	38	70.4	875	12	BF979563	BF979563	602288510	930	38	70.4	1132	13	BM544932	BM544932	AGENCOURT
858	38	70.4	878	12	BF101681	BF101681	601753411	931	38	70.4	1146	12	BG287041	BG287041	602382337
859	38	70.4	880	9	AL532835	AL532835	AL532835	932	38	70.4	1146	14	BM805511	BM805511	AGENCOURT
860	38	70.4	880	12	BG676911	BG676911	602623522	933	38	70.4	1147	12	BF204761	BF204761	AGENCOURT
861	38	70.4	881	12	BF974708	BF974708	AL535234	934	38	70.4	1150	12	BF795008	BF795008	602256121
862	38	70.4	883	9	AL535234	AL535234	AL535234	935	38	70.4	1157	12	BG288445	BG288445	602383417
863	38	70.4	888	9	A1068824	A1068824	mga0004b	936	38	70.4	1181	10	BE283174	BE283174	601103742
864	38	70.4	889	9	AL565832	AL565832	AL565832	937	38	70.4	1181	13	BM563416	BM563416	AGENCOURT
865	38	70.4	903	14	BQ278393	BQ278393	AGENCOURT	938	38	70.4	1187	14	BQ071378	BQ071378	AGENCOURT
866	38	70.4	904	17	CNS04811	AL279298	Tetradon	939	38	70.4	1197	17	AG108900	AG108900	Pan Trogl
867	38	70.4	907	12	BE974252	BE974252	601680221	940	38	70.4	1279	14	BQ958868	BQ958868	AGENCOURT

C 941	38	70.4	1280	14	BQ220875	BQ220875	AGENCOURT
C 942	38	70.4	1315	14	BM905484	BM905484	AGENCOURT
C 943	38	70.4	1360	14	BM912153	BM912153	AGENCOURT
C 944	37.5	69.4	207	10	BM513306	BM513306	
C 945	37.5	69.4	407	9	AI592143	AI592143	mr26a06.y
C 946	37.5	69.4	410	9	AA162672	AA162672	mr26a06.t
C 947	37.5	69.4	411	17	AO871675	AO871675	nbe004k
C 948	37.5	69.4	419	10	AM286717	AM286717	LGI_204.D
C 949	37.5	69.4	443	14	BQ848530	BQ848530	OGA7H01.Y
C 950	37.5	69.4	516	17	AO915576	AO915576	nbe00059G
C 951	37.5	69.4	575	14	BQ874412	BQ874412	OG15D22.Y
C 952	37.5	69.4	585	14	BF484091	BF484091	WHE1788.C
C 953	37.5	69.4	585	14	BQ999923	BQ999923	OGG23G20.
C 954	37.5	69.4	600	12	BG800123	BG800123	2112-94.M
C 955	37.5	69.4	627	14	BQ874401	BQ874401	OG15D11.Y
C 956	37.5	69.4	654	10	AM537739	AM537739	G0119D11-
C 957	37.5	69.4	659	12	BQ605495	BQ605495	H3031G08-
C 958	37.5	69.4	665	14	BQ804579	BQ804579	WHE3556.D
C 959	37.5	69.4	668	10	BQ003268	BQ003268	
C 960	37.5	69.4	771	17	AZ133663	AZ133663	OSJNB010
C 961	37.5	69.4	791	17	AZ133637	AZ133637	OSJNB010
C 962	37.5	69.4	930	12	BF103270	BF103270	60164685
C 963	37.5	69.4	1001	13	BI950915	BI950915	HVSMEL1002
C 964	37.5	69.4	1725	11	AY107917	AY107917	Zea mays
C 965	37.5	69.4	1956	11	AK005699	AK005699	Mus muscu
C 966	37.5	69.4	1956	11	TA129E07P	TA129E07P	T. brucei
C 967	37.5	68.5	144	12	BF192174	BF192174	243371.MA
C 968	37.5	68.5	152	10	AM637611	AM637611	D160C12.W
C 969	37.5	68.5	154	10	BM580331	BM580331	
C 970	37.5	68.5	176	12	BF836669	BF836669	
C 971	37.5	68.5	178	9	AA411164	AA411164	z726d04.s
C 972	37.5	68.5	180	10	BB341631	BB341631	
C 973	37.5	68.5	183	14	BQ564778	BQ564778	9124C07.Y
C 974	37.5	68.5	197	10	AM366763	AM366763	RC6-HT014
C 975	37.5	68.5	199	12	BF822944	BF822944	MRO-RT003
C 976	37.5	68.5	210	9	AV262962	AV262962	
C 977	37.5	68.5	211	10	BB562284	BB562284	
C 978	37.5	68.5	211	12	BF839450	BF839450	RC4-HT013
C 979	37.5	68.5	214	12	BF900788	BF900788	CM2-MT019
C 980	37.5	68.5	216	12	BF039916	BF039916	BP250015A
C 981	37.5	68.5	217	13	BG982690	BG982690	IL5-CN006
C 982	37.5	68.5	230	10	BB386148	BB386148	
C 983	37.5	68.5	230	14	N65680	N65680	20720.Lambda
C 984	37.5	68.5	236	10	AV336675	AV336675	
C 985	37.5	68.5	245	12	BG363828	BG363828	dc94d01.Y
C 986	37.5	68.5	258	10	AM326428	AM326428	
C 987	37.5	68.5	261	9	AV210165	AV210165	
C 988	37.5	68.5	262	10	BM478205	BM478205	
C 989	37.5	68.5	265	10	AM478202	AM478202	18999.MAR
C 990	37.5	68.5	266	10	BB243519	BB243519	
C 991	37.5	68.5	268	10	BB307932	BB307932	
C 992	37.5	68.5	268	17	AZ579049	AZ579049	30F08.Sho
C 993	37.5	68.5	269	17	AO104548	AO104548	HS_3127.A
C 994	37.5	68.5	269	17	BH358224	BH358224	CH230-175
C 995	37.5	68.5	269	17	U49985	U49985	HSU49885.Hu
C 996	37.5	68.5	274	10	BB417629	BB417629	
C 997	37.5	68.5	274	10	BB598044	BB598044	
C 998	37.5	68.5	274	13	BM054833	BM054833	1e6h03.x
C 999	37.5	68.5	276	9	AL726601	AL726601	
C 1000	37.5	68.5	276	9	AL726601	AL726601	

## ALIGNMENTS

RESULT 1  
LOCUS AA411980 318 bp mRNA linear EST 12-AUG-1997  
DEFINITION 265503.s1 Soares\_testis\_NHT Homo sapiens CDNA clone IMAGE:727252  
ACCESSION AA411980  
VERSION AA411980.1 GI:2070632  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 318)  
Hillier, L., Allen, M., Bowles, L., Dubuque, J., Geisler, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, Y., Wyllie, T., Waterston, R., and Wilson, R.  
WashU-Merck EST project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL: contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 2072 Std Error: 0.00  
Seq primer: -41ml3 fwd. ET from Amersham  
High quality sequence stop: 306.  
location/Qualifiers  
1. 318  
/organism="Homo sapiens"  
/db\_xref="GDB:5924164"  
/db\_xref="taxon:9606"  
/clone="IMAGE:727252"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - 0190(dn) primer [5]. TGTACCAATCGACAGTGGCGCGCCGACATTTTCTTTTCTTTT 3'}. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 90 a 72 c 58 g 98 t  
ORIGIN

Alignment Scores: 246 Length: 318  
Pred. No.: 47.00 Matches: 8  
Score: 100.00% Conservative: 1  
Percent Similarity: 88.89% Mismatches: 0  
Best Local Similarity: 87.04% Indels: 0  
Query Match: 9 Gaps: 0  
DB:

US-10-008-355-25 (1-10) x AA411980 (1-318)  
Oy 1 ThrGlyGlyAsnSerGlySerProVal 9  
|||||  
Db 219 ACGGAGAGAACTCTGCACGTCCTATT 193  
|||||

RESULT 2  
LOCUS BH400391 426 bp DNA linear GSS 11-DEC-2001  
DEFINITION AG-ND-147H4.YF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4.  
ACCESSION BH400391  
VERSION BH400391.1 GI:17346607  
KEYWORDS GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.  
REFERENCE 1 (bases 1 to 426)  
Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M., and Loftus, B. J.  
Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL  
COMMENT Unpublished (2001)  
Other\_GSSs: AG-ND-147H4.TR  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@igf.org  
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For  
Class: BAC ends.

FEATURES  
source Location/Qualifiers  
1..426  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-147H4"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 111 a 79 c 76 g 160 t

ORIGIN

Alignment Scores:  
Pred. No.: 330 Length: 426  
Score: 47.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 87.04% Indels: 0  
DB: 17 Gaps: 0

US-10-008-355-25 (1-10) x BH400391 (1-426)

QY 1 ThnGlyGlyAsnSerGlySerProVal 9  
|||||  
Db 81 ACAGAGGTAACCTCAGTTCCTCAATT 55

RESULT 3  
BB622700 608 bp mRNA linear EST 26-OCT-2001  
DEFINITION BB622700 RIKEN full-length enriched, adult male olfactory brain Mus musculus cDNA clone 6430525L24 5', mRNA sequence.  
VERSION BB622700  
KEYWORDS BB622700.1 GI:16461651  
SOURCE EST  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 608)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshinoda Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222

Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome-gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanka,I., Alzawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome-gsc.riken.go.jp) for further details.  
e mouse tissues.

FEATURES  
source Location/Qualifiers  
1..608  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="6430525L24"  
/clone\_lib="RIKEN full-length enriched, adult male olfactory brain"  
/sex="male"  
/tissue\_type="olfactory brain"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15', GAGAGAGAGCGCCGCACTCGAGTTTCTTTTCTTTT 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAGATTCTGAGTTATTTAATTAATTCCTCCCTCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 170 a 112 c 145 g 181 t

ORIGIN

Alignment Scores:  
Pred. No.: 473 Length: 608  
Score: 47.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 80.00% Mismatches: 0  
Query Match: 87.04% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x BB622700 (1-608)

QY 1 ThnGlyGlyAsnSerGlySerProValp10  
|||||  
Db 561 TCTGGTGAATTCCTGAAACCCAGTCTTT 590



TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR4&f2=MR4-RT0045-250401-009-906&t3=2001-04-25&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 14  
Location/Qualifiers  
1. 497  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RT0045"  
/dev\_stage="Adult"  
/note="Organ: Kidney; tumor; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 54 a 164 c 176 g 102 t 1 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 818 Length: 497  
Score: 45.00 Matches: 8  
Percent Similarity: 90.00% Conservative: 1  
Best Local Similarity: 80.00% Mismatches: 1  
Query Match: 83.33% Indels: 0  
DB: 14 Gaps: 0  
US-10-008-355-25 (1-10) x BQ327884 (1-497)  
QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
Db 81 ACGGTGGCGAGCAGGTAGCCCTGCTTC 110  
RESULT 7  
BG714149/c 796 bp mRNA linear EST 08-MAY-2001  
LOCUS 602674594F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4797094 5',  
DEFINITION mRNA sequence.  
ACCESSION BG714149  
VERSION BG714149.1 GI:13993080  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 796)  
NIH-MGC http://mhc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraaki Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10682 row: 1 column: 23  
High quality sequence stop: 677.  
Location/Qualifiers  
1. 796  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4797094"  
/clone\_lib="NIH\_MGC\_96"  
/issue\_type="hypothalamus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site: 1: BamHI; Site: 2: SalI-XhoI (gtcgag)  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 172 a 226 c 275 g 123 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,32e+03 Length: 796  
Score: 45.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 83.33% Indels: 0  
DB: 12 Gaps: 0  
US-10-008-355-25 (1-10) x BG714149 (1-796)  
QY 1 ThrGlyGlyAsnSerGlySerProVal 9  
Db 770 ACAGGTGGCAACAGCGTGACCTGTG 744  
RESULT 8  
BF128636 914 bp mRNA linear EST 24-OCT-2000  
LOCUS 601810895F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4054078 5',  
DEFINITION mRNA sequence.  
ACCESSION BF128636  
VERSION BF128636.1 GI:10967676  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 914)  
NIH-MGC http://mhc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Louis M. Straudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM893 row: m column: 23  
High quality sequence stop: 647.  
Location/Qualifiers  
1. 914  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4054078"  
/clone\_lib="NIH\_MGC\_48"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 257 a 237 c 266 g 154 t

ORIGIN

Alignment Scores:

Pred. No.:	1.51e+03	Length:	914
Score:	45.00	Matches:	8
Percent Similarity:	90.00%	Conservative:	1
Best Local Similarity:	80.00%	Mismatches:	1
Query Match:	83.33%	Indels:	0
DB:	12	Gaps:	0

US-10-008-355-25 (1-10) x BF128636 (1-914)

Oy 1 ThcglyasnseryglyserProvalphe 10  
|||||

Db 701 ACAGCGGCAATCGGCATCTCTTTC 730

RESULT 9  
BG115436

LOCUS 602316560F1 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:4416831 5',  
mRNA sequence.

ACCESSION BG115436

VERSION BG115436.1 GI:12608942

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1108)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps@email.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LLM10148 row: h column: 16  
High quality sequence stop: 680.

FEATURES  
source  
1.1108  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="4416831"  
/clone\_lib="NIH\_MGC\_88"  
/tissue\_type="duodenal adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: small intestine; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT 203 a 367 c 312 g 226 t

ORIGIN

Alignment Scores:

Pred. No.:	1.84e+03	Length:	1108
Score:	45.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	2

Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 83.33% Indels: 0  
DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BG115436 (1-1108)

Oy 2 GlycylasnseryglyserProvalphe 10  
|||||

Db 615 GGGGGAACATCGACATCGATTTTC 641

RESULT 10  
BF893687/c

LOCUS 237 bp mRNA linear EST 18-JAN-2001

DEFINITION OV1-MT0166-131100-476-b02 MT0166 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF893687

VERSION BF893687.1 GI:12285146

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 237)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV1&l2=QV1-MT0166-131100-476-b02&l3=2000-11-13&l4=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 182.

FEATURES  
source  
1.237  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MT0166"  
/dev\_stage="Adult"  
/note="Organ: marrow; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 40 a 60 c 52 g 85 t

ORIGIN

Alignment Scores:

Pred. No.:	565	Length:	237
Score:	44.00	Matches:	7
Percent Similarity:	90.00%	Conservative:	2
Best Local Similarity:	70.00%	Mismatches:	1
Query Match:	81.48%	Indels:	0
DB:	12	Gaps:	0

US-10-008-355-25 (1-10) x BF893687 (1-237)

Oy 1 ThcglyasnseryglyserProvalphe 10

```

Db      236 ACAGGGGGAACACTGGAAGCACTTTT 207
RESULT 11
LOCUS   AM060086/c
DEFINITION 687001G07.xl 687 - Early embryo from Delaware Zea mays cDNA, mRNA
SEQUENCE.
ACCESSION AM060086
VERSION   AM060086.1
KEYWORDS  EST.
SOURCE    Zea mays.
ORGANISM  Zea mays.
REFERENCE 1 (bases 1 to 279)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
           Department of Biological Sciences
           Stanford University
           855 California Ave, Palo Alto, CA 94304, USA
           Tel: 650 723 2227
           Fax: 650 725 8221
           Email: walbot@stanford.edu
           Plate: 687001 Row: G Column: 07.
FEATURES
  source
    1..279
    /organism="Zea mays"
    /cultivar="Illinois High Oil"
    /db_xref="taxon:4577"
    /clone_lib="687 Early embryo from Delaware"
    /tissue_type="embryo"
    /dev_stage="14, 21, 28, and 35 days after pollination"
    /lab_host="E. coli SOLR"
    /note="Organ: embryo; Vector: pBluescript SK; Site: 1: XhoI
    the Uni-ZAP XR system (Stratagene BN97328-12). Clones
    were picked by a Q-bot after blue/white selection
    (ampicillin resistance - use 100 micrograms/microliter).
    Developed from a pool of equal amounts of RNA from
    developing embryos sampled at 14, 21, 28 and 35 days after
    pollination of the Illinois High Oil Maize Strain Cycle
    90. This closed strain has been selected for high oil
    concentration for 90 generations and originates from the
    1890s era open pollinated variety Burr's White"
BASE COUNT 61 a 78 c 68 g 72 t
ORIGIN
Alignment Scores:
Pred. No.: 666 Length: 279
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 10 Gaps: 0
US-10-008-355-25 (1-10) x AM060086 (1-279)
Oy      1 ThrglyGlyAsnSerGlySerPro 8
Db      155 ACAGAGAGAAACAGTGCACGCCG 132
RESULT 12
LOCUS   BJ099916/c
DEFINITION 298 bp mRNA linear EST 12-DEC-2001
Xenopus laevis cDNA clone XL155108 3', mRNA sequence.
ACCESSION BJ099916
VERSION   BJ099916.1
KEYWORDS  GI:17602460

```

```

KEYWORDS  EST.
SOURCE    African clawed frog.
ORGANISM  Xenopus laevis
REFERENCE 1 (bases 1 to 298)
AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara
TITLE     Expressed genes in X. laevis embryo
JOURNAL   Unpublished (2001)
COMMENT   Contact: Tadasu Shin-I
           Center For Genetic Resource Information
           National Institute of Genetics
           1111 Yata, Mishima, Shizuoka 411-8540, Japan
           Tel: 81-559-81-6856
           Fax: 81-559-81-6855
           Email: tshini@genes.nig.ac.jp.
FEATURES
  source
    1..298
    /organism="Xenopus laevis"
    /db_xref="taxon:8355"
    /clone_lib="XLI55108"
    /dev_stage="stage 10.5"
    /note="Vector: pBSRN3; Site: 1: NotI; Site: 2: EcoRI. cDNAs
    were oligo-dT primed and directionally cloned. Staging
    according to Nieuwkoop and Faber. Library is subtracted
    and was constructed by N. Garrett and A.M. Zorn,
    (Wellcome/CRC Institute)."
BASE COUNT 102 a 72 c 62 g 58 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 712 Length: 298
Score: 44.00 Matches: 8
Percent Similarity: 80.00% Conservative: 0
Best Local Similarity: 80.00% Mismatches: 2
Query Match: 81.48% Indels: 0
DB: 13 Gaps: 0
US-10-008-355-25 (1-10) x BJ099916 (1-298)
Oy      1 ThrglyGlyAsnSerGlySerProValpHe 10
Db      35 ACTGGGGGCAATAGTGCAGCCCCCTTTT 6
RESULT 13
LOCUS   AA476126
DEFINITION VH17d08.r1 Soares_mammary_gland_NBMKG Mus musculus cDNA clone
ACCESSION AA476126
VERSION   AA476126.1
KEYWORDS  GI:2203977
SOURCE    EST.
ORGANISM  Mus musculus.
REFERENCE 1 (bases 1 to 403)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
           Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
           Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
           Theisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
           Waterston,R.
TITLE     The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   Contact: Marra M/Mouse EST Project
           WashU-HHMI Mouse EST Project
           Washington University School of MedicineP
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

```

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:515207  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 382.  
 Location/Qualifiers

# FEATURES

source

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:875727"  
 /clone\_lib="Soares\_mammary\_gland\_NBMKG"  
 /sex="male"  
 /tissue\_type="mammary gland"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia  
 ) with a modified polylinker; Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5',  
 TGTACCAATCTGAAGTGGAGCGCGCGCATGCTTTTCTTTTCTTTTCTTTT  
 T 3']]; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT7T3 vector.  
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library  
 constructed and normalized by Bento Soares and M. Fatima  
 Bonaldo."

BASE COUNT 115 a 61 c 86 g 141 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 965 Length: 403  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AA476126 (1-403)

OY 2 G1G1G1AysnserGlyserProvalphe 10  
 |||||:::|||||  
 Db 84 GCTGGCTCTTCGATCCCACTATT 110

## RESULT 14

W72307 420 bp mRNA linear EST 16-OCT-1996  
 LOCUS z60908.s1 Soares\_fetal\_heart\_NbH19w Homo sapiens cDNA clone  
 IMAGE:345086 3', mRNA sequence.  
 W72307  
 IMAGE:345086 3', mRNA sequence.

ACCESSION W72307  
 VERSION W72307.1 GI:1382930  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 420)  
 Hillier, L., Clark, N., Dubuque, T., Ellisdon, K., Hawkins, M., Holman,  
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston,  
 R., Williamson, A., Wohlmann, P. and Wilson, R.  
 The Mashu-Merck EST Project  
 Unpublished (1995)

## JOURNAL

COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 389 Std Error: 0.00  
 Seq primer: mob.REGA+ET  
 High quality sequence stop: 325.  
 Location/Qualifiers

# FEATURES

source

/organism="Homo sapiens"  
 /db\_xref="GDB:1270461"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:345086"  
 /clone\_lib="Soares\_fetal\_heart\_NbH19w"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a  
 modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 TGTACCAATCTGAAGTGGAGCGCGCGCATGCTTTTCTTTTCTTTTCTTTT  
 T 3']]; double-stranded cDNA was size selected, ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pT7T3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by  
 M. Fatima Bonaldo. This library was constructed from the  
 same fetus as the fetal lung library, Soares fetal lung  
 NbH19w."

BASE COUNT 129 a 121 c 88 g 79 t 3 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 1,01e+03 Length: 420  
 Score: 44.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 81.48% Indels: 0  
 DB: 14 Gaps: 0

US-10-008-355-25 (1-10) x W72307 (1-420)

OY 1 Thrg1G1G1AysnserGlyserProval 9  
 |||||  
 Db 201 ACAGGTGCACACACGCGANTCCCGTG 227

## RESULT 15

AW743569 432 bp mRNA linear EST 27-APR-2000  
 LOCUS ur23d03.y1 Soares\_mouse\_NMBP Mus musculus cDNA clone IMAGE:3025157  
 DEFINITION 5', mRNA sequence.  
 ACCESSION AW743569  
 VERSION AW743569.1 GI:7655383  
 KEYWORDS EST.  
 SOURCE house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 432)  
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CCAP).  
 Tumor Gene Index  
 Unpublished (1997)

## JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgsb@stratell.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:1053865

Seq primer: -40RP from Gibco  
 High quality sequence stop: 404.  
 Location/Qualifiers

# FEATURES

source

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3025157"  
 /clone\_lib="Soares\_mouse\_NMBP"



BASE COUNT	125 a	77 c	96 g	134 t
ORIGIN	Bonaldo."			
Alignment Scores:				
Pred. No.:	1.04e+03	Length:	432	
Score:	44.00	Matches:	8	
Percent Similarity:	100.00%	Conservative:	1	
Best Local Similarity:	88.89%	Mismatches:	0	
Query Match:	81.48%	Indels:	0	
DB:	10	Gaps:	0	
US-10-008-355-25 (1-10) x	AM743569 (1-432)			
Oy	2	GI:GlyAsnSerGlySerProValphe 10		
Db	352	GGTGGGCTCTTGAGATCCCACTATT 378		
RESULT 16				
LOCUS	AM743601	432 bp	mRNA	linear
DEFINITION	ur3h03.y1 Soares_mouse_NMBP Mus musculus cDNA clone IMAGE:3025205			
ACCESSION	AM743601			
VERSION	AM743601.1			
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	1 (bases 1 to 432)			
JOURNAL	NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).			
FEATURES	Tumor Gene Index			
source	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cga@bbs-rcmail.nih.gov			
	This clone is available royalty-free through LNL ; contact the			
	IMAGE Consortium (info@image.llnl.gov) for further information.			
	MG1:1053913			
	Seq primer: -40RP from Gibco			
	High quality sequence stop: 424.			
	Location/Qualifiers			
	1. .432			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:3025205"			
	/clone_lib="Soares_mouse_NMBP"			
	/lab_host="DH10B (phage-resistant)"			
	/note="Organ: bone (pooled); Vector: pT73D-Pac (Pharmacia)			
	1) with a modified polylinker; Site_1: NotI; Site_2: EcoRI;			
	1st strand cDNA was primed with a Not I - oligo(dT) primer			
	[5' TGTTACCAATCTGAGGAGGCGCGACGCGCTTTTTTTTTTTTTTT			
	3']; double-stranded cDNA was ligated to Eco RI adaptors			
	(Pharmacia), digested with Not I and cloned into the Not			
	I and Eco RI sites of the modified pT73 vector. Library			
	constructed and normalized by Bento Soares and M.Fatima			
	Bonaldo."			
BASE COUNT	124 a.	77 c	96 g	125 t
ORIGIN				
Alignment Scores:				
Pred. No.:	1.04e+03	Length:	432	
Score:	44.00	Matches:	8	
Percent Similarity:	100.00%	Conservative:	1	
Best Local Similarity:	88.89%	Mismatches:	0	
Query Match:	81.48%	Indels:	0	
DB:	10	Gaps:	0	

Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	81.48%	Indels:	0
DB:	10	Gaps:	0
US-10-008-355-25 (1-10) x AM743601 (1-432)			
OY	2	GlyGlyAnSerGlySerProValPhe	10
DB	352	GGTGGGCTCTTCGATCCCACTATT	378
RESULT 17			
LOCUS	AL117983		
DEFINITION	AL117983	447 bp	mRNA
ACCESSION	p9223a54	Beddington mouse dissected endoderm	Mus musculus cDNA
VERSION	clone 528_9K12 5'		mRNA sequence.
KEYWORDS	AL117983	1	GI:5919822
SOURCE	EST.		
ORGANISM	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 447)		
JOURNAL	Harrison,S.M., Dunwoodie,S.L., Atkell,R.M., Lehrach,H. and Beddington,R.S.		
MEDLINE	Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo		
COMMENT	Development 121 (8), 2479-2489 (1995)		
CONTACT	95401865		
CONTACT	Contact: Ruiz,P., Lehrach,H. and Avner,P.		
GENOSCOPE	EEC Mouse Transcript Mapping Consortium		
GENOSCOPE	Genoscope - CNS		
EMAIL	2, rue Gaston Cremieux, CP 5706, 91057 Evry CEDEX, France		
CLONE	Email: parner@pasteur.fr		
CLONE	clone available from Ressourcenzentrum, Heubnerweg 6, D-14059 Berlin, Germany. Web site http://www.rzp.de		
SEQ PRIMER	Seq primer: CCGGTCGAGATCCCGGT.		
FEATURES	Location/Qualifiers		
SOURCE	1..447		
	/organism="Mus musculus"		
	/strain="C57BL/6 x DBA"		
	/db_xref="taxon:10090"		
	/clone="528_9K12"		
	/clone_lib="Beddington mouse dissected endoderm"		
	/tissue_type="dissected endoderm"		
	/dev_stage="7.5 dpc"		
	/note="Vector: pSport1; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. - High quality sequence only submitted. - Average insert size: 1.2 kb (range: 0.2 - 2.kb)"		
BASE COUNT	132 a	80 c	97 g
ORIGIN		138 t	
Alignment Scores:			
Pred. No.:	1.07e+03	Length:	447
Score:	44.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	81.48%	Indels:	0
DB:	9	Gaps:	0
US-10-008-355-25 (1-10) x AL117983 (1-447)			
OY	2	GlyGlyAnSerGlySerProValPhe	10
DB	267	GGTGGGCTCTTCGATCCCACTATT	293
RESULT 18			
LOCUS	AA175055		
DEFINITION	AA175055	450 bp	mRNA
	ms82g01.i1	Scores mouse.3NBDS	Mus musculus cDNA
	5', mRNA sequence.		linear
			EST 16-FEB-1997
			IMAGE:618096

ACCESSION AAI75055  
 VERSION AAI75055.1 GI:1756176  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 450)  
 AUTHORS Marto,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Maria M/Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:378920  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 449.  
 Location/Qualifiers  
 1..450  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:618096"  
 /clone\_1lb="Soares mouse 3NbMS"  
 /sex="male"  
 /tissue\_type="Spleen"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; site\_1: Not I; site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCGAGCGAGCGCGCGCGCTGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 128 a 81 c 102 g 139 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.08e+03 Length: 450  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 Gaps: 0  
 US-10-008-355-25 (1-10) x AAI75055 (1-450)  
 Oy 2 GlycylasnsErglyserProvalphe 10  
 Db 339 GGTGGCTCTCTCGATCCCGAGTATT 365  
 RESULT 19  
 LOCUS ALO34929 491 bp mRNA linear EST 29-DEC-1999  
 DEFINITION m8708a54 Beddington mouse dissected endoderm Mus musculus cDNA  
 accession ALO34929  
 version ALO34929.1 GI:6646555  
 keywords EST.

SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 491)  
 AUTHORS Harrison,S.M., Dunwoodie,S.L., Arkell,R.M., Lehnach,H. and Beddington,R.S.  
 TITLE Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo  
 JOURNAL Development 121 (8), 2479-2489 (1995)  
 MEDLINE 95401865  
 COMMENT Contact: Miles,M., Lehnach,H. and Avner,P.  
 EEC Mouse Transcript Mapping Consortium  
 Genoscope - CNS  
 2, rue Gaston Cremieux, 91000 EVRY, France  
 Email: pavner@pasteur.fr  
 clone available from Ressourcenzentrum, Heubnerweg 6, D-14059 Berlin, Germany. Web site http://www.rzpd.de  
 Seq primer: CCGGTCGCGAATCCCGGT;  
 Vector: pSport1; site\_1: NotI; site\_2: SalI;  
 High quality sequence only submitted.  
 Cloned unidirectionally.  
 Dissected endoderm 7.5 days.  
 Average insert size: 1.2 kb (range: 0.2 - 2.kb).  
 Location/Qualifiers  
 1..491  
 /organism="Mus musculus"  
 /strain="C57BL/6 x DBA"  
 /db\_xref="taxon:10090"  
 /clone="528.9K12"  
 /clone\_1lb="Beddington mouse dissected endoderm"  
 /tissue\_type="dissected endoderm"  
 /dev\_stage="7.5 dpc"  
 /note="Vector: pSport1; site\_1: NotI; site\_2: SalI; Cloned unidirectionally.~ High quality sequence only submitted.~ Average insert size: 1.2 kb (range: 0.2 - 2.kb)"

BASE COUNT 150 a 75 c 106 g 158 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.18e+03 Length: 491  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 Gaps: 0  
 US-10-008-355-25 (1-10) x ALO34929 (1-491)  
 Oy 2 GlycylasnsErglyserProvalphe 10  
 Db 232 GGTGGCTCTCTCGATCCCGAGTATT 258  
 RESULT 20  
 LOCUS B1440854 528 bp mRNA linear EST 12-MAR-2002  
 DEFINITION 1c77c12.y1 Melton Normalized Mixed Mouse Pancreas 1 NI-MMS1 Mus musculus cDNA clone IMAGE:5659774 5', mRNA sequence.  
 accession B1440854  
 version B1440854.1 GI:15265544  
 keywords EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 528)  
 AUTHORS Melton,D., Brown,J., Keny,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Seearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,D., Bistaina,A., Schmitt,A., Theising,B., Rilter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarisvilli,R., Williams,T., Jackson,Y. and Bowers,Y.

**TITLE** Endocrine Pancreas Consortium  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu  
 Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For Information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu)  
 MGI:1946100 This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov  
 High quality sequence stop: 436.

**FEATURES**  
 source  
 1. 528  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="ICR"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5659774"  
 /clone\_lib="Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1"  
 /sex="Both for embryonic & newborn, male for adult and adult islet"  
 /dev\_stage="Embryonic day 10.5, E12.5, E16.5, newborn, adult, mixed"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1; Site\_1: Not I; Site\_2: Sal I; Five libraries representing E10.5/12.5 pancreatic bud, E16.5 pancreas, newborn pancreas, adult pancreas, and adult islets of Langerhans were separately constructed using Superscript plasmid library kit (Life Technologies). cDNA was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldi, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

**BASE COUNT** 162 a 83 c 110 g 172 t 1 others  
**ORIGIN**

**Alignment Scores:**  
 Pred. No.: 1.27e+03 Length: 528  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 13 Gaps: 0

US-10-008-355-25 (1-10) x B1440854 (1-528)

**QY** 2 GYGLYAsnSerGlySerProValphe 10  
 Db 36 GGTGGGTCTTGTGAGATCCCAAGTATT 62

**RESULT 21**  
 BG751121/c 715 bp mRNA linear EST 15-MAY-2001  
 LOCUS BG751121  
 DEFINITION 602729618F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4873276 5',  
 mRNA sequence.  
 ACCESSION BG751121  
 VERSION BG751121.1 GI:14061774  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

**REFERENCE**  
 1 (bases 1 to 715)  
**AUTHORS** NIH-MGC http://mgi.nci.nih.gov/  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1CM1751 row: c column: 05  
 High quality sequence stop: 159.

**FEATURES**  
 source  
 1. 715  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4873276"  
 /clone\_lib="NIH\_MGC\_43"  
 /tissue\_type="normal pigmented retinal epithelium"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: eye; Vector: pOT87; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library. I"

**BASE COUNT** 163 a 263 c 155 g 134 t  
**ORIGIN**

**Alignment Scores:**  
 Pred. No.: 1.72e+03 Length: 715  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 88.89% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BG751121 (1-715)

**QY** 2 GYGLYAsnSerGlySerProValphe 10  
 Db 486 GGTGGGAGTTCCGGTCGCGGTTTC 460

**RESULT 22**  
 BG918747 768 bp mRNA linear EST 05-JUN-2001  
 LOCUS BG918747  
 DEFINITION 602819216F1 NCI\_GCAP\_Mam6 Mus musculus cDNA clone IMAGE:4948443 5',  
 mRNA sequence.  
 ACCESSION BG918747  
 VERSION BG918747.1 GI:14299223  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

**REFERENCE**  
 1 (bases 1 to 768)  
**AUTHORS** NIH-MGC http://mgi.nci.nih.gov/  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished (1999)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-r@mail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10899 row: 0 column: 04  
 High quality sequence stop: 733.

## FEATURES

## Source

Location/Qualifiers

1..768

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:494843"

/clone\_lib="NCI CGAP Mam6"

/sex="female, virgin"

/tissue\_type="infiltrating ductal carcinoma"

/dev\_stage="5 months"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: PCMV-SPORT6; Site: 1; Salt; Site: 2; Notti: Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 229 a 206 c 203 g 130 t

## ORIGIN

## Alignment Scores:

Pred. No.: 1.85e+03 Length: 768  
 Score: 44.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 81.48% Indels: 0  
 DB: 13 Gaps: 0

US-10-008-355-25 (1-10) x BG918747 (1-768)

OY 1 ThrglyGlyAsnSerGlySerPro 8

DB 422 ACAGAGGTAATTCTGCAGCCCA 445

## RESULT 23

AO912379/c

LOCUS

DEFINITION nbe0036114f CUGI Rice BAC Library (ECORI) Oryza sativa genomic

clone nbe0036114f, DNA sequence.

ACCESSION

AO912379

VERSION

KEYWORDS

GSS.

SOURCE

ORGANISM

Oryza sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Eurharididae; Oryzaceae; Oryza.

1 (bases 1 to 812)

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

CONTACT: Wing RA

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAAATGACACTCATATAGG

Class: BAC ends

High quality sequence start: 26

High quality sequence stop: 287.

Location/Qualifiers

1..812

/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db\_xref="taxon:4530"

/clone="nbe0036114f"

/clone\_lib="CUGI Rice BAC Library (ECORI)"

/tissue\_type="leaf"

/lab\_host="E. coli DH10B"

/note="Vector: pBACindio8; Site: 1; EcorI; Site: 2; EcorI;

Rice is the most important food crop in the world. Half of

the world population, especially those inhabiting highly

populated areas of the humid tropics and subtropics, rely

on rice as their primary source of carbohydrate.

Monocotyledonous rice is a diploid plant (2n=24) with a

haploid genome equivalent of 431 Mbp (Arumuganathan and

Earle, 1991). The relatively small genome of rice, three

times larger than that of Arabidopsis, makes it suitable

for genomic studies. In order to facilitate positional

cloning, physical mapping and genome sequencing of rice,

we have constructed a BAC library from Oryza sativa.

Nipponbare variety using EcorI as the cloning enzyme. The

library contains 55,296 clones with an average insert size

of 121 kb providing approximately 15 haploid genome

equivalents. The deep coverage allows the isolation a

particular sequence with a probability of 99.9%. Three

high density filters, each containing 18,432 clones

(doubly spotted), represent the whole library for colony

screening and can be requested from the Clemson University

BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 194 a 221 c 159 g 237 t 1 others

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.96e+03 Length: 812  
 Score: 44.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 81.48% Indels: 0  
 DB: 17 Gaps: 0

US-10-008-355-25 (1-10) x AO912379 (1-812)

OY 2 GlyGlyAsnSerGlySerProValPhe 10

DB 703 GGAGGGAATCGTGTCCCGCTTTTC 677

## RESULT 24

BF668889/c

LOCUS

DEFINITION

602123292P1 NIH\_MGC\_56 Homo sapiens cDNA clone IMAGE:4280134 5,

BF668889

ACCESSION

BF668889

VERSION

KEYWORDS

EST.

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 823)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: CLONETECH Laboratories, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI106 row: h column: 23

High quality sequence stop: 516.

Location/Qualifiers

1..823

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4280134"

/clone\_lib="NIH\_MGC\_56"

/tissue\_type="primitive neuroectoderm"

/lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggccgctggcgc); Site\_2: SfiI (ggccattggcgc);  
 Double-stranded cDNA was prepared from cell line RNA. 5'  
 and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCGCGCGCGACATG-dt(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."

BASE COUNT 221 a 250 c 223 g 129 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	1.98e+03	Length:	823
Score:	44.00	Matches:	8
Percent Similarity:	88.89%	Conservative:	0
Best Local Similarity:	88.89%	Mismatches:	1
Query Match:	81.48%	Indels:	0
DB:	12	Gaps:	0

US-10-008-355-25 (1-10) x BF668889 (1-823)

QY 2 GlycylAsnSerGlySerProValPhe 10  
 |||||  
 Db 678 GGTGGCACAGCGGGCCCGGTTT 652

## RESULT 25

Bi695748 857 bp mRNA linear EST 18-SEP-2001  
 LOCUS 603347002F1 NCI\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5374416 5',  
 DEFINITION mRNA sequence.

ACCESSION Bi695748  
 VERSION Bi695748.1 GI:15658377  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 857)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LAM11953 row: d column: 01  
 High quality sequence stop: 811.

## FEATURES

source

1. .857  
 Location/Qualifiers

/organism="Mus musculus"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5374416"  
 /clone\_lib="NCI\_CGAP\_Mam2"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT  
 ORIGIN

261 a 156 c 177 g 263 t

## Alignment Scores:

Pred. No.:	2.07e+03	Length:	857
Score:	44.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	81.48%	Indels:	0
DB:	13	Gaps:	0

US-10-008-355-25 (1-10) x Bi695748 (1-857)

QY 2 GlycylAsnSerGlySerProValPhe 10  
 |||||  
 Db 48 GGTGGCTCTTCGATCCCACTATT 74

## RESULT 26

BO881134 958 bp mRNA linear EST 16-AUG-2002  
 LOCUS BO881134  
 DEFINITION AGENCOURT\_8728373 NIH\_MGC\_47 Homo sapiens cDNA clone IMAGE:6339187  
 5', mRNA sequence.

ACCESSION BO881134  
 VERSION BO881134.1 GI:22273142  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 958)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LICM2513 row: j column: 20  
 High quality sequence stop: 504.  
 Location/Qualifiers

## FEATURES

source

1. .958

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6339187"  
 /clone\_lib="NIH\_MGC\_47"  
 /tissue\_type="neuroblastoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library."

BASE COUNT 181 a 321 c 293 g 163 t  
 ORIGIN

## Alignment Scores:

Pred. No.:	2.31e+03	Length:	958
Score:	44.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	81.48%	Indels:	0
DB:	14	Gaps:	0

US-10-008-355-25 (1-10) x BO881134 (1-958)

QY 1 ThrGlyAsnSerGlySerPro 8  
 |||||

```

Db      776  ACCGGCGGGAACTCAGGCTCCCA 799

RESULT 27
LOCUS   BE794754/c                      970 bp    mRNA    linear    EST 20-SEP-2000
DEFINITION
601590584F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3944516 5',
mRNA sequence.
ACCESSION BE794754
VERSION   BE794754.1 GI:10215952
KEYWORDS EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 970)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabs-r@mail.nih.gov
          Tissue Procurement: DCTD/DRP
          CDNA Library Preparation: Ling Hong/Rubin Laboratory
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
          Plate: LLC601 row: h column: 21
          High quality sequence stop: 690.

FEATURES
Source
1..970
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3944516"
/clone_1db="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 181 a 273 c 267 g 248 t 1 others

Alignment Scores:
Pred. No.: 2.34e+03 Length: 970
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BE794754 (1-970)
Oy      1  ThrglglYAnserGlySerPro 8
Db      885  ACGGCGGCAACAGTGGTCCCC 862

RESULT 28
LOCUS   AK019499                      3200 bp    mRNA    linear    HTC 19-JAN-2002
DEFINITION
MUS musculus 0 day neonate skin CDNA, RIKEN full-length enriched
library, clone:4632417K18, homolog to CDNA: F1J22794 FIS, CLONE
K1A2442, full insert sequence.
ACCESSION AK019499
VERSION   AK019499.1 GI:12859746
KEYWORDS HTC; CAP trapper.
SOURCE    MUS musculus (Strain:C57Bl/6J) 0 day neonate skin CDNA to mRNA,
clone_1lb:RIKEN full-length enriched mouse CDNA library

ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS   Carninci, P. and Hayashizaki, Y.
TITLE     High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED    10349636
REFERENCE 2
AUTHORS   Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED    11042159
REFERENCE 3
AUTHORS   Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED    11076861
REFERENCE 4
AUTHORS   Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Knehl, P., Lewis, S., Matsuo, Y., Nkaido, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staib, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, R., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzatelli, J., Mommaerts, P., Nordone, P.,
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Welter, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.
and Hayashizaki, Y.
TITLE     Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED    11217851
REFERENCE 5 (bases 1 to 3200)
AUTHORS   Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Aizawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D.,
Schriml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
TITLE     Direct Submission
JOURNAL   Submitted (18-AUG-2000) Yoshinori Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome

```

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Riken = 10.0 and subtraction to Riken = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGATTCGAGTTAATTAATTAATTCCTCCCTCCCTCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 3 and 7 kb was selected before cloning. Vector: a modified pluscript KS(+) after bulk excision from lambda Phi I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

#### FEATURES

##### SOURCE

Location/Qualifiers  
1..3200  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="FANTOM.DB:4632417K18"  
/db\_xref="MGI:1900015"  
/db\_xref="taxon:10090"  
/clone="4632417K18"  
/tissue-type="skin"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="0 day neonate"  
261..2102  
/note="data source:SPTR, source key:09H5Y1, evidence:ISS homolog to CDNA: FLJ22794 FIS, CLONE KAITA2442 putative"  
/codon\_start=1  
/protein\_id="BAB31763.1"  
/db\_xref="GI:12859747"  
/db\_xref="MGI:1915508"  
/translation="MSCKKRRSQISFNPKNKKIKDYFSQVPEKEDNDPTVKYDSK MPDITNTRDQRLSPKTRDQDPPNKKITVTVGNASCKHMKMKELICRETSLSY AALNTLSAVREEVESQKREMLVCGKIGIYLNGPVCCIPGSLVIVIFCCCKK TQENKOFESODOSTNYVRFCIHAVSGSKRKLKCGELQKCKGLCVYGKKEITD TLRKDRGCTFIESDDMKLINDITITENPOPELDEGKLFQVAELPKNPRYSVO NSGSENRFKLEIYVNETTTLKEEKKIRAYIKESSEKRRKASLFFKHKEFGKM TRNSTPVKVKHLSRVSDSVGFLMNNNGNAGCATCFVFEKELYLLTQOHVIASTVGG ISDESMANIISQCVKTFDEELPTGDKFFPMKPFELISDKHLVAVLEKENGQV PAGLYHRIKRVPHGLIYIIGHPEGEKSIDCCVVPQSSRRKCOENFOAREPAGFC PAFSEIHFIRQSFQEMLHNSDYVTYDTSEFGSSGSPVDSNGSLVAMAAGITTCY OAGSNITIEFGSIMEISIDHMKQDKYKEMWNTISGNQNVENKLSIDF"

##### CDS

polyA\_signal  
3179..3184  
/note="putative"  
polyA\_site  
3200  
/note="putative"  
BASE COUNT 1048 a 599 c 701 g 852 t  
ORIGIN

#### Alignment Scores:

Pred. No.: 7.8e+03 Length: 3200  
Score: 44.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 81.48% Indels: 0  
DB: 11 Gaps: 0

US-10-008-355-25 (1-10) x AK019499 (1-3200)

Oy 2 glyglyasnserylgserprovalphe 10

Db 1878 GGTGGGTCTTCTGCATCCCAATATT 1904

#### RESULT 29

BAB176886

LOCUS

DEFINITION

BAB176886 RIKEN full-length enriched, adult male hypothalamus Mus

ACCESSION

BAB176886

VERSION

BAB176886.1 GI:8835969

KEYWORDS

house mouse.

SOURCE

house mouse.

ORGANISM

house mouse.

REFERENCE

house mouse.

AUTHORS

house mouse.

TITLE

house mouse.

JOURNAL

house mouse.

COMMENT

house mouse.

FEATURES

SOURCE

Location/Qualifiers  
1..278  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="A230064D12"  
/clone\_lib="RIKEN full-length enriched, adult male hypothalamus"  
/tissue-type="hypothalamus"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'-GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GACAGAGATCTCTGAGTAAATTAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda F1C1."

BASE COUNT 74 a 74 c 66 g 64 t  
ORIGIN

## Alignment Scores:

Pred. No.: 967 Length: 278  
Score: 43.00 Matches: 8  
Percent Similarity: 80.00% Conservative: 0  
Best Local Similarity: 80.00% Mismatches: 2  
Query Match: 79.63% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x BB176886 (1-278)

OY 1 ThrGlyAsnSerGlySerProValPhe 10

DB 184 ACCCTCGGAACCGAGGCTCCCGAGTTT 213

## RESULT 30

AM308476 377 bp mRNA linear EST 08-JAN-2001

LOCUS 3608 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.

AM308476

AM308476.1 GI:6720839

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smltth@mail.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross\_match with the -m1nscore 20 and -m1match 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTTCCAGTCACGACG  
Plate: 134 row: L column: 9  
Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers  
1..377  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; site\_1: NotI; site\_2: SalI; library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 63 a 126 c 106 g 82 t

ORIGIN

Alignment Scores:

Pred. No.: 1.31e+03 Length: 377

Score: 43.00 Matches: 7

Percent Similarity: 100.00% Conservative: 2

Best Local Similarity: 79.63% Indels: 0

Query Match: 79.63% Gaps: 0

DB: 10

US-10-008-355-25 (1-10) x BB176886 (1-278)

OY 1 ThrGlyAsnSerGlySerProValPhe 10

DB 184 ACCCTCGGAACCGAGGCTCCCGAGTTT 213

RESULT 30

AM308476 377 bp mRNA linear EST 08-JAN-2001

LOCUS 3608 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.

AM308476

AM308476.1 GI:6720839

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smltth@mail.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross\_match with the -m1nscore 20 and -m1match 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTTCCAGTCACGACG  
Plate: 134 row: L column: 9  
Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers  
1..377  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; site\_1: NotI; site\_2: SalI; library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 63 a 126 c 106 g 82 t

ORIGIN

Alignment Scores:

Pred. No.: 1.31e+03 Length: 377

Score: 43.00 Matches: 7

Percent Similarity: 100.00% Conservative: 2

Best Local Similarity: 79.63% Indels: 0

Query Match: 79.63% Gaps: 0

Best Local Similarity: 77.78% Mismatches: 0  
Query Match: 79.63% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x AM308476 (1-377)

OY 2 GlyGlyAsnSerGlySerProValPhe 10

DB 292 GCGGGAACCTCAGGCAACCCGTTTAC 318

## RESULT 31

AM315936 392 bp mRNA linear EST 08-JAN-2001

LOCUS 13723 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.

AM315936

AM315936.1 GI:6745192

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smltth@mail.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross\_match with the -m1nscore 20 and -m1match 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTTCCAGTCACGACG  
Plate: 137 row: H column: 15  
Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers  
1..392  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; site\_1: NotI; site\_2: SalI; library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 54 a 133 c 118 g 87 t

ORIGIN

Alignment Scores:

Pred. No.: 1.37e+03 Length: 392

Score: 43.00 Matches: 7

Percent Similarity: 100.00% Conservative: 2

Best Local Similarity: 77.78% Mismatches: 0

Query Match: 79.63% Indels: 0

DB: 10

US-10-008-355-25 (1-10) x AM315936 (1-392)

OY 2 GlyGlyAsnSerGlySerProValPhe 10

DB 314 GCGGGAACCTCAGGCAACCCGTTTAC 340

RESULT 32

BC834820 414 bp mRNA linear EST 25-MAY-2001

LOCUS 353172 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.

DEFINITION 353172 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.



ACCESSION BG834820  
 VERSION BG834820.1 GI:14200414  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa

REFERENCE 1  
 AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 Fahrnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
 and Keele,J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt-trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.  
 PCR primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCCTCCAGTCACGACG  
 Plate: 115 row: B column: 9  
 Seq primer: ATTAGTGACACTATAG.

FEATURES  
 source  
 1. 414  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 1P1G"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="vector: pcMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 Library made from pooled tissue from day 11, 13, 15, 20,  
 and 30 embryos."

BASE COUNT 67 a 142 c 109 g 96 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.44e+03 Length: 414  
 Score: 43.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 79.63% Indels: 0  
 DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BG834820 (1-414)

OY 2 GYGYASnsSerglySerProvalphe 10  
 |||||||||||||||:|||||:  
 Db 266 GCGGGAACCTCAGGCACCCCTTTAC 292

RESULT 33 422 bp mRNA linear EST 04-OCT-2001  
 BI830377  
 LOCUS 603073426P1 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5165142 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI830377  
 VERSION BI830377.1 GI:15941927  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1  
 AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLM11410 row: d column: 07  
 High quality sequence stop: 411.

FEATURES  
 source  
 1. 422  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:5165142"  
 /clone\_lib="NIH\_MGC\_119"  
 /tissue\_type="medulla"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pcMV-SPORT6; Site\_1: NotI;  
 Site\_2: EcoRV (destroyed); RNA source normal medulla from  
 anonymous male age 27. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.3 kb, insert size range  
 0.9-3 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 013. Note:  
 this is a NIH-MGC Library."

BASE COUNT 107 a 115 c 109 g 91 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.47e+03 Length: 422  
 Score: 43.00 Matches: 7  
 Percent Similarity: 88.89% Conservative: 1  
 Best Local Similarity: 77.78% Mismatches: 1  
 Query Match: 79.63% Indels: 0  
 DB: 13 Gaps: 0

US-10-008-355-25 (1-10) x BI830377 (1-422)

OY 2 GYGYASnsSerglySerProvalphe 10  
 ||||||||||||||| |||:||||  
 Db 359 GCGGGAATTCGCGCTTCATTTC 385

RESULT 34 473 bp mRNA linear EST 12-MAR-2001  
 BG383001  
 LOCUS 3008777-MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.  
 DEFINITION BG383001.1 GI:13307473  
 ACCESSION  
 VERSION  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa

REFERENCE 1  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 Fahrnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
 and Keele,J.W.

TITLE Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt-trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -mismatch 12 options.  
 PCR primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCCTCCAGTCACGACG

Plate: 6 row: E column: 18  
 Seq primer: ATTAGTGACACTATAG.  
 Location/Qualifiers  
 1. 473

FEATURES  
 source  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 1P1G"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 Library made from pooled tissue from day 11, 13, 15, 20,  
 and 30 embryos."  
 BASE COUNT 83 a 154 c 123 g 113 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.65e+03 Length: 473  
 Score: 43.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 79.63% Indels: 0  
 DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BG835001 (1-473)

QY 2 Glycylasnserylglyserprovalphe 10  
 Db 16 GGCGGAACCTGACGACCCGCTTAC 42

RESULT 35  
 BJI90590 483 bp mRNA linear EST 24-JAN-2002  
 LOCUS BJI90590 normalized full length cDNA library, chironomata,  
 caulonemata and rhizoid-like protonemata Physcomitrella patens  
 subsp. patens cDNA clone pphn10p19 5', mRNA sequence.  
 BJI90590  
 BJI90590.1 GI:18358531

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST.  
 Physcomitrella patens subsp. patens.  
 Physcomitrella patens subsp. patens  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.  
 1 (bases 1 to 483)  
 Fujita,T., Shin-I,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,  
 Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,T. and Hasebe  
 ,M.

REFERENCE  
 AUTHORS  
 TITLE  
 Comparison of the moss Physcomitrella patens genome with flowering  
 plants genome  
 JOURNAL  
 COMMENT  
 Unpublished (2002)  
 Contact: Tadasu Shin-I  
 Center for Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855

Email: tsuhien@nig.ac.jp  
 A backbone of the vector is Bluescript II, that was in vivo  
 digested from a modified lps phage vector (Mo Bi Tec, Germany). XhoI  
 digested-5' end of cDNA is ligated to SalI site of the vector, and  
 the BamHI digested-3' end including poly-A tail is ligated to BamHI  
 site of the vector. cDNA insert could be amplified with  
 conventional T7 and T3 primers. This normalized full-length cDNA  
 library was generated basically according to the method described  
 in Genome Research 10, 1617-1630 (2000). Carninci, P. et al.  
 Protonemata were blended by the POLYTRON, and then cultivated on  
 the BCD medium containing 1um NAA (naphthalene acetic acid) for 8  
 to 11 days under the continuous light.

FEATURES  
 source  
 1. 483  
 /organism="Physcomitrella patens subsp. patens"  
 /db\_xref="taxon:145481"  
 /clone\_lib="pphn10p19"  
 /clone\_lib="normalized full length cDNA library,"

chloronemata, caulonemata and rhizoid-like protonemata"  
 /tissue\_type="mixture of chloronemata, caulonemata and  
 rhizoid-like protonemata"  
 BASE COUNT 115 a 96 c 139 g 133 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.69e+03 Length: 483  
 Score: 43.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 79.63% Indels: 0  
 DB: 13 Gaps: 0

US-10-008-355-25 (1-10) x BJI90590 (1-483)

QY 2 Glycylasnserylglyserprovalphe 10  
 Db 168 GGAGTAATTCGTCGCGCAGTCTT 194

RESULT 36  
 BG834531 584 bp mRNA linear EST 25-MAY-2001  
 LOCUS BG834531  
 DEFINITION 352788 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION BG834531  
 VERSION BG834531.1 GI:14199878  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST.  
 Sus scrofa  
 pig.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 1 (bases 1 to 584)  
 Fahrénkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
 Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
 and Keeler,J.W.  
 Design and use of two pooled tissue normalized cDNA libraries for  
 EST discovery in swine  
 Unpublished (2000)

TITLE  
 JOURNAL  
 COMMENT  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390

Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and alt-trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -match 12 options.  
 PCR primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCACGTCACGACG  
 Plate: 114 row: B column: 9  
 Seq primer: ATTAGTGACACTATAG.  
 Location/Qualifiers

FEATURES  
 source  
 1. 584  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 1P1G"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 Library made from pooled tissue from day 11, 13, 15, 20,  
 and 30 embryos."  
 BASE COUNT 90 a 193 c 168 g 133 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.04e+03 Length: 584  
 Score: 43.00 Matches: 7  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 77.78% Mismatches: 0  
 Query Match: 79.63% Indels: 0  
 DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BG834531 (1-584)

OY 2 GLYGLYAsnSerGlySerProValPhe 10  
 |||||||  
 DB 318 GGCGGAGACTGAGCAACCCGTTTAC 344

RESULT 37  
 B1545222  
 LOCUS 60318734.F1 NIH\_MGC\_95 Homo sapiens cDNA clone IMAGE:5258866 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1545222  
 VERSION B1545222  
 KEYWORDS B1545222.1 GI:15432534  
 SOURCE EST.  
 ORGANISM human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1466)  
 AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM11652 row: e column: 11  
 High quality sequence stop: 71  
 Location/Qualifiers  
 1. 1466  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5258866"  
 /clone\_lib="NIH\_MGC\_95"  
 /tissue\_type="hippocampus"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: Bluescript (modified  
 pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gccccg  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTNN-3',  
 size-selected for average insert size 2.5 kb and  
 normalized to ROP 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NHGRI/NHGRI, National  
 Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 394 a 348 c 349 g 375 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5.17e+03 Length: 1466  
 Score: 43.00 Matches: 7  
 Percent Similarity: 88.89% Conservative: 1  
 Best Local Similarity: 77.78% Mismatches: 1  
 Query Match: 79.63% Indels: 0  
 DB: 13 Gaps: 0

US-10-008-355-25 (1-10) x B1545222 (1-1466)

OY 1 ThrGlyGlyAsnSerGlySerProVal 9  
 |||||||  
 DB 1198 ACTGCTGGAATGTGTGAGATCTCTATC 1224

RESULT 38  
 A1608160/c  
 LOCUS A1608160 185 bp mRNA linear EST 21-APR-1999  
 DEFINITION val4909.y1 Scores mouse lymph node NbMLN Mus musculus cDNA clone

IMAGE:722944 5', mRNA sequence.  
 A1608160  
 VERSION A1608160.1 GI:4617327  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Scurionathli; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 185)  
 AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
 B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rittler  
 E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
 Waterston,R. and Wilson,R.  
 The WashU-NCI Mouse EST Project 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:448440  
 This read is a RESEQUENCE of a previously sequenced mouse clone  
 correct orientation)  
 Seq primer: -40RP from Glibco  
 High quality sequence stop: 184.  
 Location/Qualifiers  
 1. 185  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:722944"  
 /clone\_lib="Scores mouse lymph node NbMLN"  
 /sex="male"  
 /tissue\_type="lymph node"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: lymph node; Vector: pT7T3D-Pac (Pharmacia)  
 with a modified polylinker; Site: 1: Not I; Site 2: Eco RI;  
 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 15',  
 TGTTCACATCTGAGTGGAGCGCCGCGATCTTTTCTTTTCTTTTCTTTTCTTTT  
 3'; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. RNA  
 provided by Dr. Bertrand Jordan. Library constructed and  
 normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 37 a 47 c 52 g 49 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 934 Length: 185  
 Score: 42.00 Matches: 8  
 Percent Similarity: 88.89% Conservative: 0  
 Best Local Similarity: 88.89% Mismatches: 1  
 Query Match: 77.78% Indels: 0  
 DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x A1608160 (1-185)

OY 1 ThrGlyGlyAsnSerGlySerProVal 9  
 |||||||  
 DB 92 ACAGGAGAGAGACTGAGGCGCCCTGTC 66

RESULT 39  
 B1783638  
 LOCUS B1783638 192 bp mRNA linear EST 26-SEP-2001  
 DEFINITION kh35h01.y1 Ascaris suum male head pAMP1 v2 Chiapelli1 McCarlier  
 Ascaris suum cDNA 5', mRNA sequence.

ACCESSION B1783638  
 VERSION B1783638.1 GI:15786530  
 KEYWORDS EST.  
 SOURCE pig roundworm.  
 ORGANISM Ascaris suum.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea  
 ; Ascaridae; Ascaris.  
 1 (bases 1 to 192)  
 McCarter,J., Clifton,S., Chiappelli,B., Pape,D., Martin,J., Wylie,T.,  
 Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,  
 Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R.,  
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe  
 ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,  
 Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and  
 Wilson,R.  
 The Washington Univ. Nematode EST Project, 1999  
 Unpublished (1999)  
 Contact: McCarter JP  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 The library was constructed by Brandi Chiappelli and Dr. James  
 McCarter at Washington University, St. Louis. The cDNA was made by  
 using Dynabead oligo-dT priming (Dyna1). PCR based library using a  
 modified protocol from the SMART PCR cDNA Synthesis Kit from  
 Clontech. Directionally cloned into the UDG sites of PAMPI.  
 Dissected nematode tissues were provided by Dr. Alan Scott  
 (ascot@hshp.edu) of the School of Public Hygiene and Public Health  
 at John Hopkins University in Baltimore, MD.  
 Seq primer: -40KP from Gibco  
 High quality sequence stop: 186.  
 Location/Qualifiers  
 1..192  
 /organism="Ascaris suum"  
 /db\_xref="taxon:6253"  
 /clone\_lib="Ascaris suum male head PAMPI v2 Chiappelli  
 McCarter"  
 /sex="Male"  
 /tissue\_type="Head"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B"  
 /note="Vector: PAMPI (Gibco); Site\_1: NotI; Site\_2: SalI;  
 The library was constructed by Brandi Chiappelli and Dr.  
 James McCarter at Washington University, St. Louis. The  
 cDNA was made by using Dynabead oligo-dT priming (Dyna1).  
 PCR based library using a modified protocol from the SMART  
 PCR cDNA Synthesis Kit from Clontech. Directionally cloned  
 into the UDG sites of PAMPI. Dissected nematode tissues  
 were provided by Dr. Alan Scott (ascot@hshp.edu) of the  
 School of Public Hygiene and Public Health at John Hopkins  
 University in Baltimore, MD."

BASE COUNT 41 a 34 c 48 g 69 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 969 Length: 192  
 Score: 42.00 Matches: 8  
 Percent Similarity: 80.00% Conservative: 0  
 Best Local Similarity: 80.00% Mismatches: 2  
 Query Match: 77.78% Indels: 0  
 DB: 13 Gaps: 0

US-10-008-355-25 (1-10) x B1783638 (1-192)  
 QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
 Db 26 ACTGGAGAAATGACGGGTCCCGCTTTTC 55

RESULT 40  
 AA507304

LOCUS AA507304 301 bp mRNA linear EST 18-AUG-1997  
 DEFINITION nm49d02.s1 NCI\_CGAP\_Pt5 Homo sapiens cDNA clone IMAGE:955683  
 similar to SW:U8R\_CHICK P23913 LAMIN B RECEPTOR. ;, mRNA sequence.  
 ACCESSION AA507304  
 VERSION AA507304.1 GI:2243743  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 301)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,  
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: David B. Kitzman, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bdrrp/image/image.html  
 Insert Length: 1136 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 288.  
 Location/Qualifiers  
 1..301  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:955683"  
 /clone\_lib="NCI\_CGAP\_Pt5"  
 /sex="male"  
 /tissue\_type="prostate"  
 /lab\_host="DH10B"  
 /note="Vector: PAMPI0; mRNA made from normal prostatic  
 epithelial cells; cDNA made by oligo-dT priming.  
 Non-directionally cloned. Size-selected on agarose gel,  
 average insert size 600 bp."

BASE COUNT 54 a 89 c 81 g 77 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.53e+03 Length: 301  
 Score: 42.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 3  
 Best Local Similarity: 66.67% Mismatches: 0  
 Query Match: 77.78% Indels: 0  
 DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AA507304 (1-301)  
 QY 2 GlyGlyAsnSerGlySerProValPhe 10  
 Db 73 GGGGGAACCTGACGCAATCGATTTC 99

RESULT 41  
 AA596016 319 bp mRNA linear EST 26-SEP-1997  
 DEFINITION nm66g01.s1 NCI\_CGAP\_Lar1 Homo sapiens cDNA clone IMAGE:1088880 3'  
 similar to gb:Y00345.cdsl POLYADENYLATE-BINDING PROTEIN (HUMAN);,  
 mRNA sequence.  
 ACCESSION AA596016  
 VERSION AA596016.1 GI:2411362  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 319)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.



```

/organism="Homo sapiens"
/db_xref="GDB:5981452"
/db_xref="taxon:9606"
/clone="IMAGE:783834"
/clone_lib="Soares_total_fetus_ND2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAAGGAGCGCGCTTAATTTTATTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      53 a      108 c      100 g      91 t
ORIGIN

Alignment Scores:
Pred. No.:      1.79e+03      Length:      352
Score:          42.00         Matches:      6
Percent Similarity: 100.00%      Conservative: 3
Best Local Similarity: 66.67%      Mismatches:  0
Query Match:    77.78%         Indels:      0
DB:             9            Gaps:        0

US-10-008-355-25 (1-10) x AAA47045 (1-352)

Oy      2  GYGLYASnsrGlySerProValPhe 10
         |||||
DB      161 GGGGGAACTCAGGCATCCGATTAC 187

RESULT 44
BM716334      369 bp      mRNA      linear      EST 28-FEB-2002
LOCUS      BM716334
DEFINITION      UI-E-C11-age-e-14-0-UI_r2 UI-E-C11 Homo sapiens cDNA clone.
ACCESSION      UI-E-C11-age-e-14-0-UI 5', mRNA sequence.
VERSION      BM716334.1 GI:19029592
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 369)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
COMMENT      Contact: Soares, MB
               Program for Rat Gene Discovery and Mapping
               University of Iowa
               451 Eckstein Medical Research Building Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: msoares@blue.weeg.uiowa.edu
               Tissue Procurement: Dr. Gregg Hageman
               CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
               CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
               DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
               Clone Distribution: Researchers may obtain clones from Research
               Genetics (www.resgen.com).
               Seq primer: M13 Reverse.
FEATURES
   source      Location/Qualifiers
               1..369
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="UI-E-C11-age-e-14-0-UI"
               /clone_lib="UI-E-C11"
               /tissue_type="RPE and Choroid"

```

```

/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker. Site_1: EcoR I; Site_2: Not I;
UI-E-C11 is a normalized cDNA library containing the
following tissue(s): RPE and Choroid. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is ACCGA.
This library was created for the program, Gene Discovery
in the Visual System, supported by National Eye Institute
(NET)."

BASE COUNT      61 a      112 c      101 g      95 t
ORIGIN

Alignment Scores:
Pred. No.:      1.87e+03      Length:      369
Score:          42.00         Matches:      6
Percent Similarity: 100.00%      Conservative: 3
Best Local Similarity: 66.67%      Mismatches:  0
Query Match:    77.78%         Indels:      0
DB:             14            Gaps:        0

US-10-008-355-25 (1-10) x BM716334 (1-369)

Oy      2  GYGLYASnsrGlySerProValPhe 10
         |||||
DB      225 GGGGGAACTCAGGCATCCGATTAC 251

RESULT 45
BF083176      375 bp      mRNA      linear      EST 18-OCT-2000
LOCUS      BF083176
DEFINITION      QV0-CS0012-150900-395-f04 CS0012 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BF083176
VERSION      BF083176.1 GI:10877006
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 375)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?rl=6t2-QV0-CS0012-150
               900-395-f04&t3=2000-09-15&t4=1)
               Seq primer: puc 18 forward
               High quality sequence stop: 375.
FEATURES
   source      Location/Qualifiers
               1..375

```



## Alignment Scores:

Pred. No.: 1.91e+03 Length: 376  
 Score: 42.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 3  
 Best Local Similarity: 66.67% Mismatches: 0  
 Query Match: 77.78% Indels: 0  
 DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BF083177 (1-376)

OY 2 GlycylasnserylglyserProvalphe 10

Db 250 GGGGGAACTGAGCAATCGATTTC 276

RESULT 48

AA564287/c 383 bp mRNA linear EST 04-SEP-1997  
 LOCUS n105c11.s1 NCI\_CGAP\_Pr21 Homo sapiens CDNA clone IMAGE:985460 3',  
 DEFINITION mRNA sequence.

ACCESSION AA564287  
 VERSION AA564287  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 383)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:

www.bio.lnl.gov/bdrip/image/image.html

Insert Length: 670 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 369.  
 Location/Qualifiers

## FEATURES

## source

1..383  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="985460"  
 /clone\_lib="NCI\_CGAP\_Pr21"  
 /sex="male"  
 /tissue\_type="normal prostate"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker: 1st strand CDNA was prepared  
 from normal prostate bulk tissue, and was then primed with  
 a Not I - oligo(dT) primer. Double-stranded CDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT73 vector. Library is not normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 94 a 112 c 94 g 82 t 1 others

## ORIGIN

## Alignment Scores:

Pred. No.: 1.95e+03 Length: 383  
 Score: 42.00 Matches: 7  
 Percent Similarity: 88.89% Conservative: 1  
 Best Local Similarity: 77.78% Mismatches: 1  
 Query Match: 77.78% Indels: 0  
 DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AA564287 (1-383)

OY 2 GlycylasnserylglyserProvalphe 10  
 Db 265 GGGGGAAACGGGGGACCTGTGTTT 239

RESULT 49

AA069844 389 bp mRNA linear EST 23-DEC-1997  
 LOCUS zml5c70.f1 Stratagene pancreas (#937208) Homo sapiens CDNA clone  
 DEFINITION IMAGE:525714 5' similar to TR:G438639 G438639 LAMIN B RECEPTOR. [1]  
 ; mRNA sequence.

ACCESSION AA069844  
 VERSION AA069844  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 389)  
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chappell, B.,  
 Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W., Hawkins  
 'M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore  
 'B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,  
 Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevisan, E.,  
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Maira, M.  
 Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)

97044478  
 TITLE JOURNAL MEDLINE  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

WARNING: There is evidence that suggests that the 384-well parent  
 plate of this clone contains both human and mouse derived clones.  
 Thus, the origin of this clone is uncertain. This caution should be  
 kept in mind should you use this clone.

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 1320 Std Error: 0.00  
 Seq primer: -28m13 rev2 from Amersham

High quality sequence stop: 182.  
 Location/Qualifiers

## FEATURES

## source

1..389  
 /organism="Homo sapiens"  
 /db\_xref="GDB:3917143"  
 /db\_xref="taxon:9606"  
 /clone\_image="525714"  
 /clone\_lib="Stratagene pancreas (#937208)"  
 /lab\_host="SOLR cells (kanamycin resistant)"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site: 1:  
 EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:  
 Oligo dT. Pancreatic adenocarcinoma cell line. Average  
 insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
 sequence: 5' GAATTCGACGACGAG 3' -3' adaptor sequence: 5'  
 CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 65 a 116 c 114 g 89 t 5 others

## ORIGIN

## Alignment Scores:

Pred. No.: 1.98e+03 Length: 389  
 Score: 42.00 Matches: 6  
 Percent Similarity: 100.00% Conservative: 3  
 Best Local Similarity: 66.67% Mismatches: 0  
 Query Match: 77.78% Indels: 0  
 DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AA069844 (1-389)

OY 2 GlycylasnserylglyserProvalphe 10

|||||







GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:41:33 ; Search time 0.536692 Seconds  
(without alignments)  
548.227 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TCGNSGSPVF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues  
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	54	100.0	213 3 US-08-523-373-22	Sequence 22, Appl
2	54	100.0	214 3 US-08-523-373-23	Sequence 23, Appl
3	54	100.0	215 3 US-08-523-373-24	Sequence 24, Appl
4	54	100.0	344 1 US-08-657-192-3	Sequence 3, Appl
5	54	100.0	344 3 US-08-523-373-5	Sequence 5, Appl
6	54	100.0	392 3 US-08-523-373-6	Sequence 6, Appl
7	54	100.0	532 1 US-08-657-192-9	Sequence 9, Appl
8	54	100.0	532 3 US-08-523-373-7	Sequence 7, Appl
9	54	100.0	537 1 US-08-657-192-15	Sequence 15, Appl
10	49	90.7	311 4 US-09-134-001C-3729	Sequence 3729, Ap
11	48	88.9	15 2 US-08-332-562A-67	Sequence 67, Appl
12	44	81.5	252 4 US-09-071-035-428	Sequence 428, App
13	44	81.5	284 4 US-09-071-035-426	Sequence 426, App
14	43	79.6	222 1 US-08-090-048-1	Sequence 1, Appl
15	43	79.6	222 2 US-08-292-550-1	Sequence 1, Appl
16	43	79.6	222 2 US-07-927-661A-1	Sequence 1, Appl
17	42	77.8	418 4 US-09-342-653-7	Sequence 7, Appl
18	39	72.2	379 1 US-07-723-002C-6	Sequence 6, Appl
19	36	66.7	379 2 US-08-887-365-36	Sequence 36, Appl
20	36	66.7	393 4 US-09-194-905-13	Sequence 13, Appl
21	36	66.7	628 2 US-07-952-853-22	Sequence 22, Appl
22	35	64.8	628 4 US-08-914-848-22	Sequence 22, Appl
23	35	64.8	8 4 US-08-296-791-7	Sequence 7, Appl
24	35	64.8	8 4 US-08-296-791-8	Sequence 8, Appl
25	35	64.8	8 5 PCT-US95-10661A-7	Sequence 7, Appl
26	35	64.8	8 5 PCT-US95-10661A-8	Sequence 8, Appl
27	35	64.8	318 4 US-09-060-756-727	Sequence 727, App

28	35	64.8	334 4 US-09-060-756-728	Sequence 728, App
29	35	64.8	503 1 US-08-245-294-8	Sequence 8, Appl
30	35	64.8	503 1 US-08-474-499-8	Sequence 8, Appl
31	35	64.8	503 1 US-08-307-279A-8	Sequence 8, Appl
32	35	64.8	503 4 US-09-525-310-8	Sequence 8, Appl
33	35	64.8	503 5 PCT-US95-06211-8	Sequence 8, Appl
34	35	64.8	550 4 US-08-868-373-8	Sequence 8, Appl
35	35	64.8	551 2 US-09-033-537A-1	Sequence 1, Appl
36	35	64.8	1394 4 US-08-296-791-2	Sequence 2, Appl
37	35	64.8	1394 5 PCT-US95-10661A-2	Sequence 2, Appl
38	35	64.8	1507 6 5268270-2	Patent No. 5268270
39	35	64.8	1541 4 US-08-296-791-3	Sequence 3, Appl
40	35	64.8	1541 5 PCT-US95-10661A-3	Sequence 3, Appl
41	35	64.8	1545 4 US-08-296-791-4	Sequence 4, Appl
42	35	64.8	1545 5 PCT-US95-10661A-4	Sequence 4, Appl
43	35	64.8	1702 4 US-08-296-791-5	Sequence 5, Appl
44	35	64.8	1702 5 PCT-US95-10661A-5	Sequence 5, Appl
45	35	64.8	1848 4 US-08-296-791-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1  
US-08-523-373-22  
; Sequence 22, Application US/08523373  
; Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohnuye, Kazuhito  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 213 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-22  
Query Match 100.0%; Score 54; DB 3; Length 213;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10  
|||||  
Db 165 TCGNSGSPVF 174

## RESULT 2

US-08-523-373-23  
Sequence 23, Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhito  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 214 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-23

Query Match 100.0%; Score 54; DB 3; Length 214;  
Best Local Similarity 100.0%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;

QY 1 TCGNSGSPVF 10  
|||||  
Db 165 TCGNSGSPVF 174

RESULT 3  
US-08-523-373-24  
Sequence 24, Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhito  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 215 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-24

Query Match 100.0%; Score 54; DB 3; Length 215;  
Best Local Similarity 100.0%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;

QY 1 TCGNSGSPVF 10  
|||||  
Db 165 TCGNSGSPVF 174

RESULT 4  
US-08-657-192-3  
Sequence 3, Application US/08657192  
Patent No. 5747321  
GENERAL INFORMATION:  
APPLICANT: YABUTA, Masayuki  
APPLICANT: OHSUYE, Kazuhito  
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/657,192  
FILING DATE: 03-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-170086

FILING DATE: 02-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-264  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-657-192-3

Query Match 100.0%; Score 54; DB 1; Length 344;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGNSSGPVF 10  
|||||  
Db 289 TGNSSGPVF 298

RESULT 5  
US-08-523-373-5  
Sequence 5, Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhiko  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE DATA:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-5

Query Match 100.0%; Score 54; DB 3; Length 344;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGNSSGPVF 10  
|||||  
Db 289 TGNSSGPVF 298

RESULT 6  
US-08-523-373-6  
Sequence 6, Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhiko  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE DATA:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 392 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-6

Query Match 100.0%; Score 54; DB 3; Length 392;  
Best Local Similarity 100.0%; Pred. No. 0.26;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGNSSGPVF 10  
|||||  
Db 289 TGNSSGPVF 298

RESULT 7  
US-08-657-192-9  
Sequence 9, Application US/08657192  
Patent No. 5747321  
GENERAL INFORMATION:  
APPLICANT: YABUTA, Masayuki  
APPLICANT: OHSUYE, Kazuhiko  
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8

TITLE OF INVENTION: PROTEASES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/657,192  
FILING DATE: 03-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-170086  
FILING DATE: 02-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-264  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-657-192-9

Query Match 100.0%; Score 54; DB 1; Length 532;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 1 TGNSSGPVF 10  
|||||  
Db 289 TGNSSGPVF 298

RESULT 8  
US-08-523-373-7  
Sequence 7, Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhiro  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595

FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-7

Query Match 100.0%; Score 54; DB 3; Length 532;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 1 TGNSSGPVF 10  
|||||  
Db 289 TGNSSGPVF 298

RESULT 9  
US-08-657-192-15  
Sequence 15, Application US/08657192  
Patent No. 5747321  
GENERAL INFORMATION:  
APPLICANT: YABUTA, Masayuki  
APPLICANT: OHSUYE, Kazuhiro  
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/657,192  
FILING DATE: 03-JUN-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-170086  
FILING DATE: 02-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-264  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 537 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-657-192-15

Query Match 100.0%; Score 54; DB 1; Length 537;  
Best Local Similarity 100.0%; Pred. No. 0.35;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGPVF 10  
DB 289 TGNSSGPVF 298

RESULT 10  
US-09-134-001C-3729  
Sequence 3729, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3729  
LENGTH: 311  
TYPE: PR  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3729

Query Match 90.7%; Score 49; DB 4; Length 311;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGNSSGPVF 10  
DB 261 GGNSSGPVF 269

RESULT 11  
US-08-332-562A-67  
Sequence 67, Application US/08332562A  
Patent No. 5985599  
GENERAL INFORMATION:  
APPLICANT: MCKENZIE, Ian F.C.  
APPLICANT: HOGARTH, Mark P.  
APPLICANT: HIBBS, Margaret L.  
APPLICANT: SCOTT, Bernadette M.  
APPLICANT: BONADONNA, Lisa  
APPLICANT: HULETT, Mark D.  
TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN  
NUMBER OF SEQUENCES: 136  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/332,562A  
FILING DATE: 31-Oct-1994  
CLASSIFICATION: A35  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/896,457  
FILING DATE: 27-May-1992

## ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 54270/119/GRHA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 67:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-332-562A-67

Query Match 88.9%; Score 48; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.089;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGPV 9  
DB 5 TGNSSGPV 13

RESULT 12  
US-09-071-035-428  
Sequence 428, Application US/09071035  
Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 466  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brooks  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 428:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 252 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-428

Query Match 81.5%; Score 44; DB 4; Length 252;  
Best Local Similarity 70.0%; Pred. No. 6.4;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNSSGPVF 10

Not 10 residues

Db 192 TGGSGSPVF 201

RESULT 13  
US-09-071-035-426  
Sequence 426, Application US/09071035  
Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8512  
TELEFAX: (301) 309-8504  
INFORMATION FOR SEQ ID NO: 426:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 284 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-426

Query Match 81.5%; Score 44; DB 4; Length 284;  
Best Local Similarity 70.0%; Pred. No. 7.2;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCNSGSPVF 10  
Db 224 TGCNSGSPVF 233

RESULT 14  
US-08-090-048-1  
Sequence 1, Application US/08090048  
Patent No. 5523237  
GENERAL INFORMATION:  
APPLICANT: Budtz, Peter M.  
APPLICANT: Nielsen, Per M.  
TITLE OF INVENTION: PROTEIN PREPARATIONS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 55232370 No. 5523237disk of No. 5523237th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/090,048  
FILING DATE: 16-Jul-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 199/91  
FILING DATE: 06-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK92/00036  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Agtis, Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3396, 214-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-090-048-1

Query Match 79.6%; Score 43; DB 1; Length 222;  
Best Local Similarity 88.9%; Pred. No. 8.1;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GNSGSPVF 10  
Db 164 GNSGSPVF 172

RESULT 15  
US-08-292-550-1  
Sequence 1, Application US/08292550  
Patent No. 5863573  
GENERAL INFORMATION:  
APPLICANT: Dammann, Claus  
APPLICANT: Budtz, Peter  
APPLICANT: Mortensen, Steen B.  
TITLE OF INVENTION: A PROCESS FOR PRODUCING CHEESE  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 58635730 No. 5863573disk of No. 5863573th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (ERO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,550  
FILING DATE: 18 AUG 1994  
CLASSIFICATION: 426  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/927,624  
FILING DATE: 31 AUG 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/0068  
FILING DATE: 8 MAR 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta A.



REGISTRATION NUMBER: 35,127  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-292-550-1

Query Match 79.6%; Score 43; DB 2; Length 222;  
Best Local Similarity 88.9%; Pred. No. 8.1;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGNGSPVF 10.  
Db 164 GGNGSPVF 172

Search completed: May 16, 2003, 13:49:04  
Job time : 1.53669 secs



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:16:27 ; Search time 1.50055 Seconds  
(without alignments)  
888.013 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TGGNSGSPVF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	10	23	AAO15228
2	54	100.0	52	23	AAO15206
3	54	100.0	52	23	AAO15207
4	54	100.0	213	17	AAK91042
5	54	100.0	214	17	AAK91043
6	54	100.0	215	17	AAK91044
7	54	100.0	336	13	AAK29644
8	54	100.0	344	17	AAK91033
9	54	100.0	344	18	AAW22218
10	54	100.0	357	13	AAK26842

11	54	100.0	392	17	AAK91034	Beta-galactosidase
12	54	100.0	532	17	AAK91035	Recombinant V8 pro
13	54	100.0	532	18	AAW22219	Protein encoded by
14	54	100.0	537	18	AAW22220	Protein encoded by
15	54	100.0	712	23	AAO15205	Porphyromonas ging
16	49	90.7	282	22	AAK61635	S. epidermidis ope
17	49	90.7	311	23	ABP38884	Staphylococcus epi
18	48	88.9	9	23	AAO15229	Porphyromonas ging
19	44	81.5	252	20	AAV00223	Enterococcus faeca
20	44	81.5	252	23	ABP43442	E faecalis EF10 a
21	44	81.5	284	20	AAV00222	Enterococcus faeca
22	44	81.5	284	23	ABP43441	E faecalis EF10 p
23	44	81.5	336	20	AAW63338	Kidney injury asso
24	43	79.6	221	22	AAE00187	Bacillus lichenifo
25	43	79.6	222	12	AAK41159	Serine protease SP
26	43	79.6	222	13	AAK26467	Sequence of a seri
27	43	79.6	222	22	AAE00031	Bacillus lichenifo
28	43	79.6	222	22	AAE00032	Bacillus lichenifo
29	43	79.6	222	22	AAE00033	Bacillus lichenifo
30	43	79.6	222	22	AAE00035	Bacillus lichenifo
31	43	79.6	222	22	AAE00036	Bacillus lichenifo
32	43	79.6	222	22	AAE00037	Bacillus lichenifo
33	43	79.6	222	22	AAE00038	Bacillus lichenifo
34	43	79.6	222	22	AAE00039	Bacillus lichenifo
35	43	79.6	222	22	AAE00040	Bacillus lichenifo
36	43	79.6	222	22	AAE00041	Bacillus lichenifo
37	43	79.6	222	22	AAE00042	Bacillus lichenifo
38	43	79.6	222	22	AAE00043	Bacillus lichenifo
39	43	79.6	222	22	AAE00044	Bacillus lichenifo
40	43	79.6	222	22	AAE00045	Bacillus lichenifo
41	43	79.6	222	22	AAE00046	Bacillus lichenifo
42	43	79.6	222	22	AAE00047	Bacillus lichenifo
43	43	79.6	222	22	AAE00048	Bacillus lichenifo
44	43	79.6	222	22	AAE00049	Bacillus lichenifo
45	43	79.6	222	22	AAE00050	Bacillus lichenifo

## ALIGNMENTS

RESULT 1	
AAO15228	
ID	AAO15228 standard; Peptide: 10 AA.
AC	AAO15228;
DT	05-SEP-2002 (first entry)
DE	Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) active site 1.
XX	
KW	Dipeptidylpeptidase-7; DPP-7; enzyme: amidolytic cleavage: active site;
KW	DPP-7 inhibitor identification: periodontal disease; gingivitis;
KW	periodontitis.
OS	Porphyromonas gingivalis.
XX	
PN	WO200238742-A2.
XX	
PD	16-MAY-2002.
XX	
PF	08-NOV-2001; 2001WO-US46782.
XX	
PR	08-NOV-2000; 2000US-246827P.
XX	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.
XX	
PI	Travis J, Potempa JS, Banbula A, Bugno M;
XX	
DR	WPI; 2002-490075/52.
XX	
PT	Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT	the dipeptidylpeptidase for protecting an animal from periodontal
PT	disease caused by Porphyromonas gingivalis

XX Claim 5: Page 32; 65pp; English.  
PS  
XX The invention comprises the amino acid and coding sequence of the  
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
CC second and third amino acids from the N-terminal end of a target peptide.  
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
CC substituent on the alpha-carbon atom of the second amino acid from the  
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
CC reducing the growth of a bacterium and protecting an animal from a  
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
CC or periodontitis). The present amino acid sequence represents an active  
CC enzyme.  
CC  
SQ Sequence 10 AA:  
Query Match 100.0%; Score 54; DB 23; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.031;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCNSGSPVF 10  
Db 1 TGCNSGSPVF 10  
IIIIIIIIII  
RESULT 2  
AA015206  
ID AA015206 standard; Protein: 52 AA.  
XX  
AC AA015206;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.  
XX  
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
XX DPP-7 inhibitor identification; periodontal disease; gingivitis;  
XX periodontitis.  
XX  
OS Porphyromonas gingivalis.  
XX  
PN W0200238742-A2.  
XX  
PD 16-MAY-2002.  
XX  
PF 08-NOV-2001; 2001WO-US46782.  
XX  
PR 08-NOV-2000; 2000US-246827P.  
XX  
PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
PI Travis J, Potempa JS, Banbula A, Bugno M;  
XX  
DR WPI: 2002-490075/52.  
XX  
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
XX the dipeptidylpeptidase for protecting an animal from periodontal  
XX disease caused by Porphyromonas gingivalis  
XX  
XX Example 6; Fig 5; 65pp; English.  
XX  
XX The invention comprises the amino acid and coding sequence of the  
XX Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
XX enzyme has amidolytic activity for cleavage of a peptide bond between the  
XX second and third amino acids from the N-terminal end of a target peptide.  
XX The DPP-7 target peptide has an aliphatic or aromatic residue as a  
XX substituent on the alpha-carbon atom of the second amino acid from the  
XX N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
XX useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
XX reducing the growth of a bacterium and protecting an animal from a

CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
CC or periodontitis). The present amino acid sequence represents the  
CC C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7  
CC (DPP-7) enzyme.  
CC  
SQ Sequence 52 AA:  
Query Match 100.0%; Score 54; DB 23; Length 52;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCNSGSPVF 10  
Db 1 TGCNSGSPVF 10  
IIIIIIIIII  
RESULT 3  
AA015207  
ID AA015207 standard; Protein: 52 AA.  
XX  
AC AA015207;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Staphylococcus aureus V8 endopeptidase C-terminal region.  
XX  
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
XX DPP-7 inhibitor identification; periodontal disease; gingivitis;  
XX periodontitis; V8 endopeptidase.  
XX  
OS Staphylococcus aureus.  
XX  
PN W0200238742-A2.  
XX  
PD 16-MAY-2002.  
XX  
PF 08-NOV-2001; 2001WO-US46782.  
XX  
PR 08-NOV-2000; 2000US-246827P.  
XX  
PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
PI Travis J, Potempa JS, Banbula A, Bugno M;  
XX  
DR WPI: 2002-490075/52.  
XX  
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
XX the dipeptidylpeptidase for protecting an animal from periodontal  
XX disease caused by Porphyromonas gingivalis  
XX  
XX Example 6; Fig 5; 65pp; English.  
XX  
XX The invention comprises the amino acid and coding sequence of the  
XX Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
XX enzyme has amidolytic activity for cleavage of a peptide bond between the  
XX second and third amino acids from the N-terminal end of a target peptide.  
XX The DPP-7 target peptide has an aliphatic or aromatic residue as a  
XX substituent on the alpha-carbon atom of the second amino acid from the  
XX N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
XX useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
XX reducing the growth of a bacterium and protecting an animal from a  
XX periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
XX or periodontitis). The present amino acid sequence represents the  
XX C-terminal region of the Staphylococcus aureus V8 endopeptidase enzyme.  
XX  
SQ Sequence 52 AA:  
Query Match 100.0%; Score 54; DB 23; Length 52;  
Best Local Similarity 100.0%; Pred. No. 0.17;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCNSGSPVF 10  
IIIIIIIIII

Db 1 TGNNGSPVF 10

RESULT 4  
AAR91042  
ID AAR91042 standard; Peptide: 213 AA.  
XX AAR91042;  
XX 23-MAY-1996 (first entry)  
XX V8 mature protease (aal-213).  
DE  
XX Linker peptide; V8 protease; Staphylococcus aureus;  
XX recombinant protein; fusion protein; beta-galactosidase;  
KW Escherichia coli; transposon Tn903;  
KM aminoglycoside 3'-phosphotransferase.  
XX  
XX Staphylococcus aureus strain V8 (ATCC 27733).  
OS  
XX EP700995-A2.  
PN 13-MAR-1996.  
PD  
XX 06-SEP-1995; 95EP-0306235.  
PF 07-NOV-1994; 94JP-0296028.  
PR 07-SEP-1994; 94JP-0238595.  
XX  
XX (SUNR ) SUNTORY LTD.  
PA  
XX Ohsuye K, Yabuta M;  
PI WPI; 1996-141021/15.  
DR  
XX Prod. of recombinant polypeptide(s) - using host cells transformed  
PT with a gene coding for the desired polypeptide fused to a protective  
PT polypeptide  
XX  
XX Example 8; Fig 14a; 44pp; English.  
PS  
XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A  
CC and C are protective polypeptides, B is Staphylococcus aureus mature  
CC V8 protease lacking the C-terminal repeat region, and L is a linker  
CC (AAR91032). It was produced as an inclusion body in Escherichia coli  
CC host cells, and was cleaved using the E. coli OmpT protease to  
CC yield active V8 protease. Extension of the C-terminal end of the  
CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)  
CC i.e. up to residue Phe-215, also resulted in the formation of  
CC inclusion bodies in E. coli. Any further extension gave a soluble  
CC product which exhibited protease activity that repressed growth of  
CC the host cells.  
CC  
XX  
XX  
SQ Sequence 213 AA;  
Query Match 100.0%; Score 54; DB 17; Length 213;  
Best Local Similarity 100.0%; Pred. No. 0.73;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGNNGSPVF 10  
Db 165 TGNNGSPVF 174

RESULT 5  
AAR91043  
ID AAR91043 standard; Peptide: 214 AA.  
XX AAR91043;  
XX 23-MAY-1996 (first entry)  
DT V8 mature protease (aal-214).  
XX  
DE

XX Linker peptide; V8 protease; Staphylococcus aureus;  
KW recombinant protein; fusion protein; beta-galactosidase;  
KW Escherichia coli; transposon Tn903;  
KM aminoglycoside 3'-phosphotransferase.  
XX  
XX Staphylococcus aureus strain V8 (ATCC 27733).  
OS  
XX EP700995-A2.  
PN 13-MAR-1996.  
PD  
XX 06-SEP-1995; 95EP-0306235.  
PF 07-NOV-1994; 94JP-0296028.  
PR 07-SEP-1994; 94JP-0238595.  
XX  
XX (SUNR ) SUNTORY LTD.  
PA  
XX Ohsuye K, Yabuta M;  
PI WPI; 1996-141021/15.  
DR  
XX Prod. of recombinant polypeptide(s) - using host cells transformed  
PT with a gene coding for the desired polypeptide fused to a protective  
PT polypeptide  
XX  
XX Example 8; Fig 14b; 44pp; English.  
PS  
XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A  
CC and C are protective polypeptides, B is Staphylococcus aureus mature  
CC V8 protease lacking the C-terminal repeat region, and L is a linker  
CC (AAR91032). It was produced as an inclusion body in Escherichia coli  
CC host cells, and was cleaved using the E. coli OmpT protease to  
CC yield active V8 protease. Extension of the C-terminal end of the  
CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)  
CC i.e. up to residue Phe-215, also resulted in the formation of  
CC inclusion bodies in E. coli. Any further extension gave a soluble  
CC product which exhibited protease activity that repressed growth of  
CC the host cells.  
CC  
XX  
XX  
SQ Sequence 214 AA;  
Query Match 100.0%; Score 54; DB 17; Length 214;  
Best Local Similarity 100.0%; Pred. No. 0.73;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGNNGSPVF 10  
Db 165 TGNNGSPVF 174

RESULT 6  
AAR91044  
ID AAR91044 standard; Peptide: 215 AA.  
XX AAR91044;  
XX 23-MAY-1996 (first entry)  
DT V8 mature protease (aal-215).  
DE  
XX Linker peptide; V8 protease; Staphylococcus aureus;  
KW recombinant protein; fusion protein; beta-galactosidase;  
KW Escherichia coli; transposon Tn903;  
KM aminoglycoside 3'-phosphotransferase.  
XX  
XX Staphylococcus aureus strain V8 (ATCC 27733).  
OS  
XX EP700995-A2.  
PN 13-MAR-1996.  
PD  
XX

PF 06-SEP-1995; 95EP-0306235.  
 XX  
 PR 07-NOV-1994; 94JP-0296028.  
 PR 07-SEP-1994; 94JP-0238595.  
 XX  
 PA (SUNR ) SUNTORY LTD.  
 XX  
 PI Ohnuye K, Yabuta M;  
 XX  
 DR WPI: 1996-141021/15.  
 XX  
 XX Prod. of recombinant polypeptide(s) - using host cells transformed  
 PT with a gene coding for the desired polypeptide fused to a protective  
 PT polypeptide  
 PS  
 XX Example 8; Fig 14c; 44pp; English.  
 CC A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A  
 CC and C are protective polypeptides, B is Staphylococcus aureus mature  
 CC V8 protease lacking the C-terminal repeat region, and L is a linker  
 CC (AAR91032). It was produced as an inclusion body in Escherichia coli  
 CC host cells, and was cleaved using the E. coli OmpT protease to  
 CC yield active V8 protease. Extension of the C-terminal end of the  
 CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)  
 CC i.e. up to residue Phe-215, also resulted in the formation of  
 CC inclusion bodies in E. coli. Any further extension gave a soluble  
 CC product which exhibited protease activity that repressed growth of  
 CC the host cells.

Sequence 215 AA:

Query Match 100.0%; Score 54; DB 17; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 0.74;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGNSGSPVF 10  
 |||||  
 DB 165 TGGNSGSPVF 174

#### RESULT 7

AAR29644  
 ID AAR29644 standard; Protein: 336 AA.

XX AAR29644;

DT 11-FEB-1993 (first entry)

XX Protease from S. Aureus.

DE Protease; PCR; amplify; Staphylococcus.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Peptide 1..68

FT Protein /label= signal\_peptide

FT 69..336

FT /label= Protease

XX JP04211370-A.

XX 03-AUG-1992.

XX 19-FEB-1991; 91JP-0024633.

XX 20-FEB-1990; 90JP-0040396.

XX (SHIO ) SHIONOGI & CO LTD.  
 XX WPI: 1992-304938/37.  
 XX N-PSDB; AAQ27988.  
 XX

PT Novel protease prep'd. using Bacillus or Saccharomyces host -  
 PT capable of cleaving peptide bond at carboxyl terminus of glutamic  
 PT acid residues in polypeptide(s)  
 XX  
 XX Disclosure: Page 15-16; 25pp; Japanese.

XX The sequences given in AAR26842 and AAR29644 are proteases which were  
 CC isolated from Staphylococcus aureus strains. The DNA sequences  
 CC encoding these proteins were isolated by PCR using the primer  
 CC sequences given in AAQ27960-86. The protease specifically cleaves  
 CC the peptide bond at the C-terminus of the glutamic acid residue in  
 CC polypeptide.

Sequence 336 AA:

Query Match 100.0%; Score 54; DB 13; Length 336;  
 Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGNSGSPVF 10  
 |||||  
 DB 233 TGGNSGSPVF 242

#### RESULT 8

AAR91033  
 ID AAR91033 standard; Protein: 344 AA.

XX AAR91033;

DT 23-MAY-1996. (first entry)

XX Beta-galactosidase-V8 protease fusion protein.

DE V8 protease; Staphylococcus aureus; recombinant protein;

KW fusion protein; beta-galactosidase; Escherichia coli.

XX Chimeric Escherichia coli;

OS Chimeric Staphylococcus aureus strain V8 (ATCC 27733).

XX Key

FT Region 1..124

FT /note= "beta-galactosidase region"

FT Region 125..344

FT /note= "mature V8 protease without the repeat  
 region"

XX EP700995-A2.

XX 13-MAR-1996.

XX 06-SEP-1995; 95EP-0306235.

XX 07-NOV-1994; 94JP-0296028.

XX 07-SEP-1994; 94JP-0238595.

XX (SUNR ) SUNTORY LTD.

XX Ohnuye K, Yabuta M;

XX WPI: 1996-141021/15.

XX Prod. of recombinant polypeptide(s) - using host cells transformed

XX with a gene coding for the desired polypeptide fused to a protective

XX polypeptide  
 XX Example 2; Page 12-13; 44pp; English.

XX Fusion proteins (AAR91033 and AAR91034) were constructed comprising  
 CC an Escherichia coli beta-galactosidase derivative (protective  
 CC polypeptide) fused at its C-terminal end to the S. aureus mature  
 CC V8 protease without or with the repeat region. The constructs  
 CC were inserted into vector pG97S4DHCT(G)R6, yielding pV8RPT(-) and

CC pV8RPT(+), respectively. Both constructs yielded active protease  
 CC when expressed in *E. coli* JM101 transformants.  
 XX  
 SQ Sequence 344 AA;  
 Query Match 100.0%; Score 54; DB 17; Length 344;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGCNGSPVF 10  
 |||||||||  
 Db 289 TGCNGSPVF 298  
 RESULT 9  
 AAW22218  
 ID AAW22218 standard; Protein: 344 AA.  
 XX  
 AC AAW22218;  
 XX  
 DT 11-SEP-1997 (first entry)  
 XX  
 DE Protein encoded by pV8RPT(-) construct.  
 XX  
 KW Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;  
 KM truncation; wild type; PCR; polymerase chain reaction; amplification;  
 XX proteolytic activity; fusion protein; beta-galactosidase; urea.  
 OS Chimeric - *Escherichia coli*.  
 OS Chimeric - Staphylococcus aureus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..100  
 FT /note="E. coli beta-galactosidase portion"  
 FT Region 101..120  
 FT /note="synthetic R6 linker"  
 FT Region 125..344  
 FT /note="truncated S. aureus V8 protease portion"  
 EP745669-A2.  
 PD 04-DEC-1996.  
 XX  
 PF 31-MAY-1996; 96EP-0303939.  
 XX  
 PR 02-JUN-1995; 95JP-0170086.  
 XX  
 PA (SUNR ) SUNTORY LTD.  
 XX  
 PI Ohsuye K, Yabuta M;  
 XX  
 DR WPI; 1997-013693/02.  
 XX  
 PT Staphylococcus aureus V8 protease mutants - with increased  
 PT resistance to denaturation  
 XX  
 PS Claim 2; Page 13-14; 42pp; English.  
 XX  
 CC The invention relates to new mutant Staphylococcus aureus V8 proteases  
 CC which have enzyme activity even under environmental conditions which  
 CC promote protein denaturation. The mutants are based on 3 truncated V8  
 CC proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino  
 CC acids from the C-terminal of the wild type protease. The mutants also  
 CC contain amino acid substitutions, especially D44E, N71S and/or R147K.  
 CC The protein sequence shown here represents a chimeric protein  
 CC comprising a truncated Staphylococcus aureus V8 protease lacking the  
 CC prepro and C-terminal 48 amino acids linked, via a synthetic linker,  
 CC downstream of the *E. coli* beta-galactosidase. The S. aureus portion of  
 CC the chimera was amplified by the primers AAR7324-5 from wild type  
 CC sequence. The coding sequence was then used to generate mutants of the  
 CC V8 protease which retain their levels of activity in the presence of a  
 CC higher concentration of protein denaturant e.g. 5 M urea.

SQ Sequence 344 AA;  
 Query Match 100.0%; Score 54; DB 18; Length 344;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGCNGSPVF 10  
 |||||||||  
 Db 289 TGCNGSPVF 298  
 RESULT 10  
 AAR26842  
 ID AAR26842 standard; Protein: 357 AA.  
 XX  
 AC AAR26842;  
 XX  
 DT 11-FEB-1993 (first entry)  
 XX  
 DE Protease from S. Aureus ATCC12600.  
 XX  
 KW Protease; PCR; amplify; Staphylococcus.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..68  
 FT /label= Signal-peptide  
 FT Protein 69..358  
 FT /label= Protease  
 XX  
 PN JP04211370-A.  
 XX  
 PD 03-AUG-1992.  
 XX  
 PF 19-FEB-1991; 91JP-0024633.  
 XX  
 PR 20-FEB-1990; 90JP-0040398.  
 XX  
 PA (SHIO ) SHIONOGI & CO LTD.  
 XX  
 DR WPI; 1992-304938/37.  
 DR N-PSDB; AAQ27987.  
 XX  
 PT Novel protease prep. using Bacillus or Saccharomyces host -  
 PT capable of cleaving peptide bond at carboxyl terminus of glutamic  
 PT acid residues in polypeptide(s)  
 XX  
 PS Disclosure; Page 13-15; 25pp; Japanese.  
 XX  
 CC The sequences given in AAR26842 and AAR29644 are proteases which were  
 CC isolated from Staphylococcus aureus strains. The DNA sequences  
 CC encoding these proteins were isolated by PCR using the primer  
 CC sequences given in AAQ27960-86. The protease specifically cleaves  
 CC the peptide bond at the C-terminus of the glutamic acid residue in  
 CC polypeptide.  
 XX  
 SQ Sequence 357 AA;  
 Query Match 100.0%; Score 54; DB 13; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGCNGSPVF 10  
 |||||||||  
 Db 233 TGCNGSPVF 242  
 RESULT 11  
 AAR91034  
 ID AAR91034 standard; Protein: 392 AA.  
 XX  
 AC AAR91034;

```

XX 23-MAY-1996 (first entry)
DT Beta-galactosidase-V8 protease fusion protein.
XX
DE
XX
XX
XX V8 protease: Staphylococcus aureus; recombinant protein:
XX fusion protein; beta-galactosidase; Escherichia coli.
OS
XX Chimeric Escherichia coli;.
XX Chimeric Staphylococcus aureus strain V8 (ATCC 27733).
XX
XX Key Location/Qualifiers
XX Region 1..124
XX /note="beta-galactosidase region"
XX Region 125..392
XX /note="mature V8 protease including the repeat
XX region"
XX
XX EP700995-A2.
XX
XX 13-MAR-1996.
XX
XX
XX 06-SEP-1995; 95EP-0306235.
XX
XX 07-NOV-1994; 94JP-0296028.
XX 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohsuye K, Yabuta M;
XX
XX WPI: 1996-141021/15.
XX
XX Prod. of recombinant polypeptide(s) - using host cells transformed
XX with a gene coding for the desired polypeptide fused to a protective
XX polypeptide
XX
XX Example 2; Page 14-15; 44pp; English.
XX
XX Fusion proteins (AAR91033 and AAR91034) were constructed comprising
XX an Escherichia coli beta-galactosidase derivative (protective
XX polypeptide) fused at its C-terminal end to the S. aureus mature
XX V8 protease without or with the repeat region. The constructs
XX were inserted into vector pG97S4Dhcr(G)R6, yielding pV8PRT(-) and
XX pV8PRT(+), respectively. Both constructs yielded active protease
XX when expressed in E. coli JM101 transformants.
XX
XX Sequence 392 AA:
SQ
Query Match 100.0%; Score 54; DB 17; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGNNGSPVF 10
| | | | | | | | | |
Db 289 TGNNGSPVF 298
| | | | | | | | | |
RESULT 12
AAR91035
ID AAR91035 standard; Protein: 532 AA.
XX
XX AAR91035;
XX
XX 23-MAY-1996 (first entry)
XX
XX Recombinant V8 protease V8D fusion protein.
XX
XX Linker peptide: V8 protease; Staphylococcus aureus;
XX recombinant protein; fusion protein; beta-galactosidase;
XX Escherichia coli; transposon Tn903;
XX aminoglycoside 3'-phosphotransferase.
XX
XX

```

```

OS Chimeric Escherichia coli;
OS Chimeric synthetic;
OS Chimeric staphylococcus aureus strain V8 (ATCC 27733);
OS Chimeric transposon Tn903.
XX
XX Key Location/Qualifiers
XX Region 1..100
XX /note="beta-galactosidase region"
XX Region 101..120
XX /note="R6 linker"
XX Region 104..105
XX /note="cleavage site for OmpT protease"
XX Region 125..335
XX /note="V8 protease region"
XX Region 336..356
XX /note="R6 linker"
XX Cleavage-site 339..340
XX /note="cleavage site for OmpT protease"
XX Region 307..532
XX /note="aminoglycoside 3'phosphotransferase
XX region"
XX
XX EP700995-A2.
XX
XX 13-MAR-1996.
XX
XX
XX 06-SEP-1995; 95EP-0306235.
XX
XX 07-NOV-1994; 94JP-0296028.
XX 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohsuye K, Yabuta M;
XX
XX WPI: 1996-141021/15.
XX
XX Prod. of recombinant polypeptide(s) - using host cells transformed
XX with a gene coding for the desired polypeptide fused to a protective
XX polypeptide
XX
XX Example 3; Page 16-18; 44pp; English.
XX
XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
XX and C are protective polypeptides (derived from Escherichia coli
XX beta-galactosidase and and Tn903 aminoglycoside 3'-phosphotransferase,
XX respectively), B is Staphylococcus aureus mature V8 protease
XX lacking the C-terminal repeat region, and L is a linker peptide
XX (AAR91032). The fusion protein is expressed in inactive form in E.
XX coli. It is then recovered, solubilised and cleaved at the linker
XX peptide regions with a protease intrinsic to the host cells, i.e.
XX OmpT protease, to allow recovery of V8 protease.
XX
XX Sequence 532 AA:
SQ
Query Match 100.0%; Score 54; DB 17; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGNNGSPVF 10
| | | | | | | | | |
Db 289 TGNNGSPVF 298
| | | | | | | | | |
RESULT 13
AAW22219
ID AAW22219 standard; Protein: 532 AA.
XX
XX AAW22219;
XX
XX 11-SEP-1997 (first entry)
XX
XX Protein encoded by pV8D construct.
XX

```



```

XX XX Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
KM truncation; wild type; PCR; polymerase chain reaction; amplification;
KM proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
OS Chimeric - Escherichia coli.
OS Chimeric - Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH Region 1..100
FT /note="E. coli beta-galactosidase portion"
FT 101..124
FT /note="R6 linker sequence"
FT 125..336
FT /note="truncated S. aureus V8 protease portion"
FT 337..360
FT /note="R6 linker sequence"
FT 361..532
FT /note="aminoglucoiside 3'-phosphotransferase portion"
FT Region
XX
XX EP745669-A2.
XX
XX 04-DEC-1996.
XX
XX 31-MAY-1996; 96EP-0303939.
XX
XX 02-JUN-1995; 95JP-0170086.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohsuye K, Yabuta M;
XX
XX WPI; 1997-013693/02.
XX
XX Staphylococcus aureus V8 protease mutants - with increased
XX resistance to denaturation
XX
XX Claim 7; Page 16-17; 42pp; English.
XX
XX The invention relates to new mutant Staphylococcus aureus V8 proteases
XX which have enzyme activity even under environmental conditions which
XX promote protein denaturation. The mutants are based on 3 truncated V8
XX proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
XX acids from the C-terminal of the wild type protease. The mutants also
XX contain amino acid substitutions, especially D44E, N71S and/or R147K.
XX The protein sequence shown here represents a chimeric protein
XX comprising a truncated Staphylococcus aureus V8 protease lacking the
XX prepro and C-terminal 56 amino acids linked, via a synthetic R6 linker,
XX downstream of the E. coli beta-galactosidase. Also included downstream of
XX the V8 protease fragment is a second R6 linker and a fragment of the
XX aminoglucoiside 3'-phosphotransferase protein. The chimeric sequence
XX was generated by restriction digestion and ligation from the V8Rpt(-)
XX sequence (see AAW22218) by using a natural EcoRV site which removed a
XX further 8 amino acid from the C-terminus. This truncated V8 protease,
XX designated V8D, retains its level of activity in the presence of a
XX higher concentration of protein denaturant e.g. 5 M urea.
XX
XX Sequence 532 AA:
XX
XX Query Match 100.0%; Score 54; DB 18; Length 532;
XX Best Local Similarity 100.0%; Pred. NO. 1.9;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGNSGSPVF 10
Db 289 TGGNSGSPVF 298
RESULT 14
AAW22220
ID AAW22220 standard; Protein: 537 AA.
XX
XX AAW22220;
AC

```

```

XX XX 11-SEP-1997 (first entry)
XX
XX Protein encoded by pV8F construct.
DE
XX
XX Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
KM truncation; wild type; PCR; polymerase chain reaction; amplification;
KM proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
XX Chimeric - Escherichia coli.
XX Chimeric - Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH Region 1..100
FT /note="E. coli beta-galactosidase"
FT 101..124
FT /note="R6 linker sequence"
FT 125..339
FT /note="truncated S. aureus V8 protease"
FT 342..365
FT /note="R6 linker sequence"
FT 366..537
FT /note="aminoglucoiside 3'-phosphotransferase"
FT Region
XX
XX EP745669-A2.
XX
XX 04-DEC-1996.
XX
XX 31-MAY-1996; 96EP-0303939.
XX
XX 02-JUN-1995; 95JP-0170086.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohsuye K, Yabuta M;
XX
XX WPI; 1997-013693/02.
XX
XX Staphylococcus aureus V8 protease mutants - with increased
XX resistance to denaturation
XX
XX Claim 8; Page 19-20; 42pp; English.
XX
XX The invention relates to new mutant Staphylococcus aureus V8 proteases
XX which have enzyme activity even under environmental conditions which
XX promote protein denaturation. The mutants are based on 3 truncated V8
XX proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
XX acids from the C-terminal of the wild type protease. The mutants also
XX contain amino acid substitutions, especially D44E, N71S and/or R147K.
XX The protein sequence shown here represents a chimeric protein
XX comprising a truncated Staphylococcus aureus V8 protease lacking the
XX prepro and C-terminal 53 amino acids linked, via a synthetic R6 linker,
XX downstream of the E. coli beta-galactosidase. Also included downstream of
XX the V8 protease fragment is a second R6 linker and a fragment of the
XX aminoglucoiside 3'-phosphotransferase protein. The chimeric sequence
XX was generated by restriction digestion and ligation from the V8Rpt(-)
XX sequence (see AAW22218) by using a natural EcoRV site which removed a
XX further 8 amino acid from the C-terminus. This truncated V8 protease,
XX designated V8F, retains its level of activity in the presence of a
XX higher concentration of protein denaturant e.g. 5 M urea.
XX
XX Sequence 537 AA:
XX
XX Query Match 100.0%; Score 54; DB 18; Length 537;
XX Best Local Similarity 100.0%; Pred. NO. 1.9;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGNSGSPVF 10
Db 289 TGGNSGSPVF 298
RESULT 15

```

AA015205  
 ID AA015205 standard; Protein: 712 AA.  
 XX  
 AC AA015205;  
 XX  
 DT 05-SEP-2002 (first entry)  
 XX  
 DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).  
 XX  
 KM Dipeptidylpeptidase-7; DPP-7; enzyme: amidolytic cleavage;  
 KM DPP-7 inhibitor identification; periodontal disease; gingivitis;  
 KM periodontitis.  
 XX  
 OS Porphyromonas gingivalis.  
 XX  
 PN WO200238742-A2.  
 PD 16-MAY-2002.  
 XX  
 PF 08-NOV-2001; 2001WO-US46782.  
 XX  
 PR 08-NOV-2000; 2000US-246827P.  
 XX  
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
 XX  
 PI Travis J, Potempa JS, Banbula A, Bugno M;  
 XX  
 DR WPI: 2002-490075/52.  
 DR N-PSDB; AAL43635.  
 XX  
 PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
 PT the dipeptidylpeptidase for protecting an animal from periodontal  
 PT disease caused by Porphyromonas gingivalis  
 XX  
 PS Claim 7; Fig 4; 65pp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequence of the  
 CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
 CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
 CC second and third amino acids from the N-terminal end of a target peptide.  
 CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
 CC substituent on the alpha-carbon atom of the second amino acid from the  
 CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
 CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
 CC reducing the growth of a bacterium and protecting an animal from a  
 CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
 CC or periodontitis). The present amino acid sequence represents the  
 CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the  
 CC invention.  
 CC  
 SQ Sequence 712 AA;  
 XX  
 Query Match 100.0%; Score 54; DB 23; Length 712;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TCGNSGSPVF 10  
 |||||||||  
 Db 644 TCGNSGSPVF 653

Search completed: May 16, 2003, 13:43:49  
 Job time : 1.50055 secs

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:17:32 ; Search time 0.383352 Seconds  
(Without alignments)  
1081.940 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TCGNCGSPVF 10

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: ~112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSPROT.40:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	336	1 STSP_STAUA	P04188 staphylococ
2	43	79.6	316	1 GSEP_BACLI	P80057 bacillus 11
3	42	77.8	418	1 ER24_HUMAN	O76062 homo sapien
4	41	75.9	521	1 YF92_MYCPN	O50335 mycoplasma
5	39	72.2	328	1 PYRD_AGRAE	P28294 agrocycbe ae
6	39	72.2	378	1 PLXB_ASPNG	O00205 aspergillus
7	39	72.2	607	1 V66K_BWYVF	P09506 beet wester
8	39	72.2	997	1 YNM3_YEAST	P53920 saccharomyc
9	39	72.2	1035	1 RRPD_BWYVF	P09507 beet wester
10	38	70.4	434	1 PYRD_ARATH	P32746 arabidopsis
11	37	68.5	401	1 HME1_MOUSE	P09065 mus musculu
12	36	66.7	320	1 CEL1_AGABI	O00023 agaricus bi
13	36	66.7	380	1 PLXA_COGL	O00374 colleotric
14	36	66.7	425	1 Y350_HAEIN	P24326 haemophilus
15	36	66.7	508	1 DEGP_RICCN	O92141 rickettsia
16	36	66.7	513	1 DEGP_RICPR	O05942 rickettsia
17	36	66.7	516	1 Y067_MYCGE	P42254 aspergillus
18	36	66.7	628	1 ABFA_ASPNG	P42254 aspergillus
19	36	66.7	921	1 ITH4_PIG	P79263 sus scrofa
20	35	64.8	108	1 Y108_NPVP	O10347 otygia pseu
21	35	64.8	417	1 IRX5_HUMAN	P78411 homo sapien
22	35	64.8	503	1 DEGP_BARHE	P54925 bartonella
23	35	64.8	504	1 DEGP_RHIME	O52894 rhizobium m
24	35	64.8	513	1 DEGP_MYCUB	O44597 brucella ab
25	35	64.8	524	1 Y395_MYCGE	P47635 mycoplasma
26	35	64.8	531	1 YF88_MYCPN	O50339 mycoplasma
27	35	64.8	533	1 YAB3_MYCPN	P75610 mycoplasma
28	35	64.8	584	1 YMB3_YEAST	O04228 saccharomyc
29	35	64.8	1394	1 HAP1_HAEIN	P45387 haemophilus
30	35	64.8	1409	1 HAP1_HAEIN	P45387 haemophilus
31	35	64.8	1532	1 IGA1_NEIGO	P09790 natsesia 9
32	35	64.8	1541	1 IGA1_HAEIN	P42782 haemophilus
33	35	64.8	1545	1 IGA3_HAEIN	P45385 haemophilus

34	35	64.8	1694	1 IGA0_HAEIN	P44969 haemophilus
35	35	64.8	1702	1 IGA2_HAEIN	P45384 haemophilus
36	35	64.8	1714	1 GSEP_DROME	P28668 drosophila
37	35	64.8	1849	1 IGA4_HAEIN	P45386 haemophilus
38	34	63.0	154	1 KRSC_CHICK	P04459 gallus gall
39	34	63.0	157	1 YF89_MYCPN	O50338 mycoplasma
40	34	63.0	303	1 Y007_MYCLE	O32870 streptococci
41	34	63.0	325	1 YF65_STREN	O97908 streptococci
42	34	63.0	353	1 YF91_MYCPN	O50336 mycoplasma
43	34	63.0	424	1 CBPT_THEYU	P29068 thermocactin
44	34	63.0	485	1 CAP1_NICPL	P49315 nicotiana p
45	34	63.0	492	1 CAT2_LYCES	O9xh3 lycopersico

## ALIGNMENTS

```

RESULT 1
STSP_STAUA STANDARD: PRT: 336 AA.
ID STSP_STAUA
AC P04188:
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutaryl endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine
DE protease) (V8 protease) (Endopeptidase Glu-C).
GN SSPA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V8.
RX MEDLINE=87316953; PubMed=3306605;
RA Carmona G., Gray G.L.;
RT "Nucleotide sequence of the serine protease gene of Staphylococcus
RT aureus, strain V8.";
RL Nucleic Acids Res. 15:6757-6757(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20569178; PubMed=11119502;
RA Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.;
RT "Description of staphylococcus serine protease (ssp) operon in
RT staphylococcus aureus and nonpolar inactivation of sspA-encoded serine
RT protease.";
RL Infect. Immun. 69:159-169(2001).
RN [3]
RP SEQUENCE OF 69-280.
RC STRAIN=V8.
RX MEDLINE=78212487; PubMed=96922;
RA Drapau G.R.;
RT "The primary structure of staphylococcal protease.";
RL Can. J. Biochem. 56:534-544(1978).
CC -1- FUNCTION: PREFERENTIALLY CLEAVES PEPTIDE BONDS ON THE CARBOXYL-
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL cleavage: Asp-I-Xaa, Glu-I-Xaa.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
CC -1- DATABASE: NAME=worthington-biochem.com/manual/p/stap.html".
CC WWW="http://www.worthington-biochem.com/manual/p/stap.html".
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y00356; CAAG8434.1; -
DR EMBL: AF309515; AAG45843.1; -
DR PIR: A26812; PRSASK.
DR MEROPS: S01.269; -
DR InterPro: IPR000126; Ser_proteas_V8.

```

```

DR InterPro: IPR001254; Ser-protase_Try.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00839; V8PROTEASE.
DR PROSITE: PS00672; V8_HIS_1.
DR PROSITE: PS00673; V8_SER_1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 29
FT PROPEP 30 68
FT CHAIN 69 336
FT ACT_SITE 119 119
FT ACT_SITE 161 161
FT ACT_SITE 237 237
FT CONFLICT 109 109
FT CONFLICT 125 125
FT CONFLICT 145 145
FT CONFLICT 193 193
FT CONFLICT 229 229
FT CONFLICT 259 261
FT CONFLICT 268 270
SQ SEQUENCE 336 AA; 36326 MW; 8B138DC7996AA3E CRC64;

Query Match          100.0%; Score 54; DB 1; Length 336;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGPVF 10
Db 233 TCGNSGPVF 242

RESULT 2
GSEP_BACLI STANDARD: PRT: 316 AA.
ID GSEP_BACLI
AC P80057;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamy1 endopeptidase precursor (EC 3.4.21.15) (Glutamate specific
  endopeptidase) (GSE).
GN BLASE.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 14580;
RX MEDLINE=93054737; PubMed=1429718;
RA Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,
  Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;
  "Purification, characterization, cloning, and expression of a
  glutamic acid-specific protease from Bacillus licheniformis ATCC
  14580.";
RT J. Biol. Chem. 267:23782-23788(1992).
RN [2]
RP SEQUENCE OF 95-316.
RX MEDLINE=92155199; PubMed=1346764;
RA Svendsen I., Bredam K.;
  "Isolation and amino acid sequence of a glutamic acid specific
  endopeptidase from Bacillus licheniformis.";
RT Eur. J. Biochem. 204:165-171(1992).
RN [3]
RP FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE
  CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES. WITH A STRONG
  PREFERENCE FOR GLU.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Asp-|-Xaa, Glu-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10060; BAA0949.1;
DR PIR: S23078; S23078.
DR PIR: A45134; A45134.
DR MEROPS: S01.271;
DR InterPro: IPR000126; Ser-protase_V8.
DR InterPro: IPR001254; Ser-protase_Try.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00020; TRYP-SPC; 1.
DR PROSITE: PS00672; V8_HIS_1.
DR PROSITE: PS00673; V8_SER_1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 30
FT PROPEP 31 94
FT CHAIN 95 316
FT ACT_SITE 141 141
FT ACT_SITE 261 261
FT DISULFID 126 142
FT DISULFID 275 279
SQ SEQUENCE 316 AA; 33611 MW; 96D7552CB7089B09 CRC64;

Query Match          79.6%; Score 43; DB 1; Length 316;
Best Local Similarity 88.9%; Pred. No. 2.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGNSSGPVF 10
Db 258 GGNSSGPVF 266

RESULT 3
ER24_HUMAN STANDARD: PRT: 418 AA.
ID ER24_HUMAN
AC O76062; Q95982; Q96621; Q96664;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Delta(14)-sterol reductase (EC 1.3.1.70) (C-14 sterol reductase)
  (Sterol C14-reductase) (Delta14-SR) (Transmembrane 7 superfamily
  member 2) (Another new gene 1) (Putative sterol reductase SR-1).
GN TM7SF2 OR ANGI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277456; PubMed=9615229;
RA Lemmens I.H., Kas K., Merregaert J., Van De Ven W.J.M.;
  "Identification and molecular characterization of TM7SF2 in the FAUNA
  gene cluster on human chromosome 11q13.";
RT Genomics 49:437-442(1998).
RN [2]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX MEDLINE=99097347; PubMed=9878250;
RA Holmer L., Pezhman A., Worman H.J.;
  "The human lamin B receptor/sterol reductase multigene family.";
RT Genomics 54:469-476(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Eye;
RA Strausberg R.;
  Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
RN [4]
RP FUNCTION.
RX PubMed=11784322;
RA Roberti R., Benatti A.M., Galli G., Caruso D., Maras B., Aisa C.,
  Becatti T., Della Fazio M.A., Servillo G.;
  "Cloning and expression of sterol Delta14-reductase from bovine
  liver.";
RT Eur. J. Biochem. 269:283-290(2002).

```

```

CC -1- FUNCTION: Involved in the conversion of lanosterol to cholesterol.
CC -1- CATALYTIC ACTIVITY: 4,4-dimethyl-5-alpha-cholesta-8,24-dien-3-
CC beta-ol + NADPH(+) = 4,4-dimethyl-5-alpha-cholesta-8,14,24-trien-3-
CC beta-ol + NADPH.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum.
CC -1- TISSUE SPECIFICITY: Expressed in adult heart, brain, pancreas,
CC lung, liver, skeletal muscle, kidney, ovary, prostate, and testis,
CC but not detected in placenta, spleen, thymus, small intestine,
CC colon (mucosal lining), or peripheral blood leukocytes.
CC -1- SIMILARITY: BELONGS TO THE ERG4/ERG24 FAMILY.
CC CAUTION: Ref.1 sequence differs from that shown due to a number of
CC sequencing problems as reported in Ref.2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF048704; AAC21457.1; ALT_FRAME.
DR EMBL: AF023676; AAC21450.1; ALT_FRAME.
DR EMBL: AF096303; AAD09769.1; -.
DR EMBL: AF096304; AAD09765.1; -.
DR EMBL: BC009052; AAH09052.1; -.
DR EMBL: BC012857; AAH12857.1; -.
DR Genbank: HGNC:11863; TM7SF2.
DR MIM: 603414; -.
DR InterPro: IPR001171; ERG4_ERG24.
DR Pfam: PF01222; ERG4_ERG24; 1.
DR PROSITE: PS01017; STEROL_REDUCT_1; 1.
DR PROSITE: PS01018; STEROL_REDUCT_2; 1.
KW Sterol biosynthesis; Oxidoreductase; NADP; Transmembrane;
KM Endoplasmic reticulum; Polymorphism.
FT TRANSMEM 13 35 POTENTIAL.
FT TRANSMEM 62 81 POTENTIAL.
FT TRANSMEM 102 124 POTENTIAL.
FT TRANSMEM 129 148 POTENTIAL.
FT TRANSMEM 255 277 POTENTIAL.
FT TRANSMEM 287 304 POTENTIAL.
FT TRANSMEM 355 377 POTENTIAL.
FT VARIANT 299 299 I->T.
FT CONFLICT 179 179 /Frid-VAR 012716.
FT SEQUENCE 418 AA; 46417 MW; 357C8ABE2BEDA918 CRC64;
SQ
Query Match 77.8%; Score 42; DB 1; Length 418;
Best Local Similarity 66.7%; Pred. No. 4.3;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 2 GGNSSGPVF 10
Db 163 GGNSSGNPLY 171

```

```

RA Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
RT "Sequence analysis of 56 kb from the genome of the bacterium
RT Mycoplasma pneumoniae comprising the data region, the atp operon and a
RT cluster of ribosomal protein genes.";
RL Nucleic Acids Res. 24:628-639(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (potential).
CC -----
CC -1- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U43738; AAC43664.1; -.
DR EMBL: AE000025; AAB95898.1; -.
DR InterPro: IPR002414; DUF30/31.
DR Pfam: PF01727; DUF30; 1.
DR Pfam: PF01732; DUF31; 1.
DR PRINTS: PR00840; Y06768FAMILY.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KM Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 521 HYPOTHETICAL LIPOPROTEIN MPN592.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 521 AA; 59500 MW; 0E706CDEC8CEBDA CRC64;
Qy 2 GGNSSGPVF 10
Db 423 GGNSSGPVF 431

```

```

RESULT 4
YF92_MYCPN STANDARD: PRT; 521 AA.
AC Q50335;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical Lipoprotein MPN592 precursor (D02_ort521).
GN MPN592 OR MP250.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=96177562; PubMed=8604303;

```

```

RESULT 5
PYRD_AGRAE STANDARD: PRT; 328 AA.
AC P28294;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)
DE (DHODHase) (DHODase) (DHOD).
GN URA1.
OS Agrocycbe aegerita.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Boletiales; Agrocycbe.
OX NCBI_TaxID=5400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9308391; PubMed=1452035;
RA Noel T., Labarete J.;
RT "Sequence of the URA1 gene encoding dihydroorotate dehydrogenase from
RT the basidiomycete fungus Agrocycbe aegerita.";
RL Gene 122:233-234(1992).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
CC H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.

```

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M80295; AAA32636.1; -
DR PIR: JN0453; JN0453.
DR InterPro: IPR001295; DHO_dh.
DR Pfam: PF01180; DHOdehase; 1.
DR TIGRPFAMS: TIGR01036; pyrd_sub2; 1.
DR PROSITE: PS00911; DHOdeHASE_1; 1.
DR PROSITE: PS00912; DHOdeHASE_2; 1.
DR Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FAD.
KM NF-BIND 288 296 FAD (NAD PART) (POTENTIAL).
FT SEQUENCE 328 AA; 35085 MW; 71FA3D2A4D57EF75 CRC64;

Query Match 72.2%; Score 39; DB 1; Length 328;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNSSGSPVF 10
Db 261 TGGISGAPLF 270

RESULT 6
ID PLYB_ASPNG STANDARD; PRT; 378 AA.
AC 000205;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Pectin lyase B precursor (EC 4.2.2.10) (PLB).
GN PLB.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 120.49 / N400;
RA MEDLINE=92357005; PubMed=1495474;
RX Kusters-Van Someren M., Filippi M., de Graaf L., van den Broeck H.,
RA Kester H., Hinnen A., Visser J.;
RT "Characterization of the Aspergillus niger pelb gene: structure and
RT regulation of expression."
RL Mol. Gen. Genet. 234:113-120(1992).
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give
CC oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-
CC enuronosyl groups.
CC -----
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65552; CAA46521.1; -
DR EMBL: A12248; CAA01023.1; -
DR HSP: Q01172; 1IDU.
DR InterPro: IPR002022; Amb_allergen.
DR Pfam: PF00544; pec_lyase; 1.
KM Lyase; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 20 OR 21 (POTENTIAL).
FT CHAIN 21 378 PECTIN LYASE B.

```

```

FT ACT_SITE 255 255 POTENTIAL.
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 378 AA; 39703 MW; 4FF321AF2B0B72FF CRC64;

Query Match 72.2%; Score 39; DB 1; Length 378;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGNSSGSPVF 10
Db 35 TGGGSASPVY 44

RESULT 7
ID V6K_BWVF STANDARD; PRT; 607 AA.
AC P09506;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last annotation update)
DE 66.2 kDa protein (ORF 2).
OS Beet western yellows virus (isolate FL-1) (BWV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Poliovirus.
OX NCBI_TaxID=12043;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=89057523; PubMed=3194229;
RA Veldt I., Lot H., Lelser M., Scheidecker D., Gullley H., Richards K.,
RA Jonard G.;
RT "Nucleotide sequence of beet western yellows virus RNA."
RL Nucleic Acids Res. 16:9917-9932(1988).
CC -1- SIMILARITY: TO POTATO LEAFROLL VIRUS ORF2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13063; CAA31463.1; -
DR PIR: S01939; S01939.
DR MEROPS: S52.001; -
DR InterPro: IPR000382; Luteo_ORF2.
DR Pfam: PF02122; Luteo_ORF2; 1.
DR PRINTS: PRD0913; LVIRUSORF2.
SQ SEQUENCE 607 AA; 66210 MW; 2A41F82911DEC84F CRC64;

Query Match 72.2%; Score 39; DB 1; Length 607;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGNSSGSPVF 10
Db 354 GGHSSGSPVF 362

RESULT 8
ID YNM3_YEAST STANDARD; PRT; 997 AA.
AC P53920;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 110.9 kDa protein in SPC98-TOM70 intergenic region.
GN YNL123W OR N1897.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97245296; PubMed=9090055;
RA de Antoni A., D'Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,
RA Pallavicini A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-13b from chromosome XIV of
RT Saccharomyces cerevisiae reveals an unusually high number of
RT overlapping open reading frames."
RL Yeast 13:261-266(1997).
CC -1- SIMILARITY: TO S.POMBE SPBC1685.05.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 269382; CAA93384.1; -
CC EMBL; 271399; CAA96004.1; -
CC MEROPS; S01.UPC; -
CC SGD; S0005067; YNL123W.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00595; PDZ; 1.
CC SMART; SM00228; PDZ; 2.
CC Hypothetical protein.
CC SEQUENCE 997 AA; 110881 MW; A26005C1DDDB932C CRC64;
SQ
Query Match 72.2%; Score 39; DB 1; Length 997;
Best Local Similarity 77.8%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGNGSPV 9
DB 232 GGGSGSPV 240

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X13063; CAA31464.2; -
CC PIR; S01940; RRVOFL.
CC InterPro; IPR001795; Luteo_RNA_pol.
CC Pfam; PF02123; Luteo_ORF3; 1.
CC PRINTS; PR00914; LVIRUSRNAPOL.
CC Transferase; RNA-directed RNA polymerase.
CC SEQUENCE 1035 AA; 115870 MW; 54642FEC88BF66F CRC64;
SQ
Query Match 72.2%; Score 39; DB 1; Length 1035;
Best Local Similarity 77.8%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGNSSSPV 10
DB 354 GGGSGSPV 362

```

```

RESULT 10
PYRD_ARATH STANDARD; PRT; 434 AA.
ID PYRD_ARATH
AC P32746;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dihydroorotate dehydrogenase, mitochondrial precursor (EC 1.3.3.1)
DE (Dihydroorotate oxidase) (DHODHase).
PYRD.
OS Arabidopsis thaliana (Mouse-ear cress).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
CC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93272056; PubMed=1303803;
RA Minet M., Dufour M.F., Lacroix F.;
RT "Complementation of Saccharomyces cerevisiae auxotrophic mutants by
RT Arabidopsis thaliana cDNAs."
RL Plant J. 2:417-422(1992).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
CC H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62909; CAA44695.1; -
CC PIR; S23762; S23762.
CC InterPro; IPR001295; DHO_dh.
CC InterPro; IPR003009; FMN_enzyme.
CC Pfam; PF01180; DHODHase; 1.
CC TIGRfam; TIGR01036; PYRD_sub2; 1.
CC PROSITE; PS00911; DHODHASE_1; 1.
CC PROSITE; PS00912; DHODHASE_2; 1.
CC Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FAD;
CC Transl. peptide; Mitochondrion.
CC TRANSIT 1
CC CHAIN ? 434 MITOCHONDRION (POTENTIAL).
CC NP_BIND 393 401 FAD (NAD PART) (POTENTIAL).
CC SEQUENCE 434 AA; 45568 MW; 186BA05F3EF49D91 CRC64;
SQ
Query Match 70.4%; Score 38; DB 1; Length 434;

```

Best Local Similarity 70.0%; Pred. No. 22;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TCGNSGSPVF 10  
||| |||  
Db 366 TGGUSGKPLF 375

## RESULT 11

HME1\_MOUSE STANDARD; PRT; 401 AA.  
AC P09065; 01-NOV-1988 (Rel. 09, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Homeobox protein engrailed-1 (Mo-En-1).  
GN EN1 OR EN-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN 11  
RP SEQUENCE FROM N.A.  
RA MEDLINE=93185339; PubMed=1363401;  
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,  
RA Provart N.J., Joyner A.L.;  
RT "Cloning and sequence comparison of the mouse, human, and chicken  
RT engrailed genes reveal potential functional domains and regulatory  
RT regions".  
RL Dev. Genet. 13:345-358(1992).  
RN 12  
RP SEQUENCE OF 278-401 FROM N.A.  
RA MEDLINE=88112776; PubMed=2892757;  
RA Joyner A.L., Martin G.R.;  
RT "En-1 and En-2, two mouse genes with sequence homology to the  
RT Drosophila engrailed gene: expression during embryogenesis".  
RL Genes Dev. 1:29-38(1987).  
RN 13  
RP SEQUENCE OF 298-401 FROM N.A.  
RA MEDLINE=86079501; PubMed=2416459;  
RA Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;  
RT "Expression during embryogenesis of a mouse gene with sequence  
RT homology to the Drosophila engrailed gene".  
RL Cell 43:29-37(1985).  
RN 14  
RP SEQUENCE OF 321-380 FROM N.A.  
RA MEDLINE=91099509; PubMed=1980115;  
RA Holland P.W.H., Williams N.A.;  
RT "Conservation of engrailed-like homeobox sequences during vertebrate  
RT evolution".  
RL FEBS Lett. 277:250-252(1990).  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: BELONGS TO THE ENGRAILED HOMEBOX FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L12703; AAA0360.2; -  
DR EMBL: Y00201; CAA68361.1; -  
DR PIR: A26629; A26629  
DR PIR: A24778; A24778  
DR PIR: S13009; S13009  
DR PIR: A48423; A48423  
DR HSSP: P02836; 3HDD.  
DR TRANSFAC: T02016; -  
DR MGD: MGI:95389; En1.  
DR InterPro: IPR00747; Engrailed.

DR InterPro: IPR001356; Homeobox.  
DR Pfam: PF00046; homeobox.1.  
DR PRINTS: PR00026; ENGRAILED.  
DR PRINTS: PR00024; HOMEBOX.  
DR Prodom: PD000010; Homeobox.1.  
DR SMART: SM00389; HOX.1.  
DR PROSITE: PS00027; HOMEBOX\_1; 1.  
DR PROSITE: PS00033; ENGRAILED; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.  
FT DOMAIN 52 87 PRO-RICH.  
FT DOMAIN 73 87 POLY-PRO.  
FT DOMAIN 207 228 POLY-ALA.  
FT DNA\_BIND 312 371 HOMEBOX.  
SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152FAE CRC64;  
Query Match 68.5%; Score 37; DB 1; Length 401;  
Best Local Similarity 75.0%; Pred. No. 31;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCGNSGSP 8  
:|||||

Db 238 SGNAGSP 245

## RESULT 12

CELL\_AGABI STANDARD; PRT; 320 AA.  
AC Q00023;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cellulose-growth-specific protein precursor.  
GN CELL.  
OS Agaricus bisporus (Common mushroom).  
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
OC Agaricales; Agaricaceae; Agaricus.  
OX NCBI\_TaxID=5341;  
RN 11  
RP SEQUENCE FROM N.A.  
RA STRAIN=D649;  
RA MEDLINE=93012985; PubMed=1398098;  
RA Raguz S., Yaguee E., Wood D.A., Thurston C.F.;  
RT "Isolation and characterization of a cellulose-growth-specific gene  
RT from Agaricus bisporus".  
RL Gene 119:183-190(1992).  
RN 12  
RP SEQUENCE FROM N.A.  
RA STRAIN=D649;  
RA MEDLINE=94237428; PubMed=8181702;  
RA Amesilla A.L., Thurston C.F., Yaguee E.;  
RT "Cell: a novel cellulose binding protein secreted by Agaricus  
RT bisporus during growth on crystalline cellulose".  
RL FEBS Microbiol. Lett. 116:293-299(1994).  
CC -1- FUNCTION: PROBABLE GLYCOSYL HYDROLASE ACTIVE ON CELLULOSE.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
CC -1- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M86356; AAA53434.1; -  
DR HSSP: P00725; 2CBH.  
DR InterPro: IPR000254; CBD\_fungal.  
DR InterPro: IPR005103; Glyco\_hydro\_61.  
DR Pfam: PF00734; CBM\_1; 1.  
DR Pfam: PF03443; Glyco\_hydro\_61; 1.



DR SMART: SM00236; fCBD; 1.  
 DR PROSITE; PS00562; CBD; FUNGAL; 1.  
 KW Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.  
 FT SIGNAL 1 29  
 FT CHAIN 30 320  
 FT DOMAIN 30 261  
 FT DOMAIN 262 285  
 FT DOMAIN 286 320  
 FT DISULFID 292 309  
 FT DISULFID 303 319  
 FT CARBOHYD 163 163  
 SO SEQUENCE 320 AA; 33754 MW; 60E2C8080895CA2B CRC64; N-LINKED (GLICNAC... ) (POTENTIAL).

Query Match  
 Best Local Similarity 66.7%; Score 36; DB 1; Length 320;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Pred. No. 36;

Oy 1 TCGNSGSP 8  
 Db 256 SGGNGSP 263

## RESULT 13

ID PLYA\_COLGL STANDARD. PRT. 380 AA.  
 AC Q00374;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Pectin lyase precursor (EC 4.2.2.10).  
 GN PNL1.  
 OS Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella  
 OS cingulata).  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 CC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;  
 CC Glomerella.  
 CC NCBI\_TaxID=5457;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 RX MEDLINE=94237480; PubMed=8181749;  
 RA Templeton M.D., Sharrock K.R., Bowen J.K., Crowhurst R.N.,  
 RA Rikerink E.H.;  
 RT "The pectin lyase-encoding gene (pn1) family from Glomerella  
 RT cingulata: characterization of pn1a and its expression in yeast.";  
 RL Gene 142:141-146(1994).  
 CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give  
 CC oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-  
 CC enuroosyl groups.  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: L22857; AAA21817.1; -  
 DR HSSP; Q01172; 11D1.  
 DR InterPro; IPR002022; Amb\_allergen.  
 DR Pfam; PF00544; pec\_lyase; 1.  
 KW Lyase; Signal; Glycoprotein.  
 FT SIGNAL 1 20  
 FT CHAIN 21 380  
 FT CARBOHYD 130 130  
 SO SEQUENCE 380 AA; 39326 MW; 3DF9A99FBB482053 CRC64; N-LINKED (GLICNAC... ) (POTENTIAL).

Query Match  
 Best Local Similarity 66.7%; Score 36; DB 1; Length 380;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Pred. No. 43;

Oy 1 TCGNSGPVF 10  
 Db 36 TCGSATPVY 45

## RESULT 14

ID Y350\_HAEIN STANDARD. PRT. 425 AA.  
 AC P24326;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein HI0350 (ORF3).  
 GN HI0350.  
 OS Haemophilus influenzae.  
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 CC NCBI\_TaxID=727;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 RP STRAIN=RM 7004 / Serotype B;  
 RX MEDLINE=92065797; PubMed=1956282;  
 RA Maskell D.J., Szabo M.J., Butler P.D., Williams A.E., Moxon E.R.;  
 RT "Molecular analysis of a complex locus from Haemophilus influenzae  
 RT involved in phase-variable lipopolysaccharide biosynthesis.";  
 RL Mol. Microbiol. 5:1013-1022(1991).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RP STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=9350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (Probable).  
 CC -1- SIMILARITY: TO E.COLI AMPG AND TO YEAST YBR220C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: X57315; CAA40569.1; -  
 DR EMBL; 032719; AAC22011.1; -  
 DR PIR; S15289; S15289.  
 DR TIGR; HI0350; -  
 DR InterPro; IPR004752; AmpC\_permease.  
 DR TIGRfams; TIGR00901; 2A0125.1.  
 KW Hypothetical protein; Transmembrane; Inner membrane; Transport;  
 KW Complete proteome.  
 FT VARIANT 5 5 L -> F (IN STRAIN RM 7004).  
 FT VARIANT 32 32 S -> L (IN STRAIN RM 7004).  
 FT VARIANT 271 271 A -> S (IN STRAIN RM 7004).  
 FT VARIANT 313 313 A -> S (IN STRAIN RM 7004).  
 FT VARIANT 415 415 L -> W (IN STRAIN RM 7004).  
 FT VARIANT 418 418 E -> K (IN STRAIN RM 7004).  
 SO SEQUENCE 425 AA; 47354 MW; 2757C3F61B08FAB5 CRC64;

Query Match  
 Best Local Similarity 66.7%; Score 36; DB 1; Length 425;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Pred. No. 49;

OY 1 TCGNSGSPVF 10  
 111111  
 Db 20 TCGNSGLPLF 29

RESULT 15  
 DEGP\_RICCN STANDARD: PRT: 508 AA.  
 ID DEGP\_RICCN 092JAL:  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable serine protease do-like precursor (EC 3.4.21.-).  
 GN DEGP OR HTRA OR RC0166.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiiae; Rickettsia.  
 OX NCBI\_TaxID=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barde V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";  
 RL Science 293:2093-2098(2001).  
 CC -I SIMILARITY: BELONGS TO PEPTIDASE FAMILY SZC.  
 CC -I SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: AE008583; AL02704.1; ALT\_INIT.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001940; Protease2C.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin; 1.  
 DR Pfam: PF00595; PDZ; 2.  
 DR PRINTS: PRO0834; PROTEASES2C.  
 DR SMART: SM00228; PDZ; 2.  
 DR PROSITE: PS50106; PDZ; 1.  
 KW Hydrolase; Serine protease; Repeat; Signal; Complete proteome.  
 FT SIGNAL 1 23  
 FT CHAIN 1 508  
 FT DOMAIN 119 284  
 FT DOMAIN 286 377  
 FT DOMAIN 413 497  
 FT ACT\_SITE 134 134  
 FT ACT\_SITE 164 164  
 FT ACT\_SITE 242 242  
 SQ SEQUENCE 508 AA: 55599 MM; DZF53A690ECD0AD7 CRC64;  
 Query Match 66.7%; Score 36; DB 1; Length 508;  
 Best Local Similarity 75.0%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 GNSGSPVF 10  
 111111  
 Db 240 GNSGSPMF 247

Search completed: May 16, 2003, 13:44:33  
 Job time : 1.38335 secs

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:38:37 ; Search time 0.733845 Seconds  
(without alignments)  
1310.011 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TGNNGSPVF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_73:\*\*  
2: PIR1:\*\*  
3: PIR2:\*\*  
4: PIR3:\*\*  
5: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	54	100.0	336	1	PRRSASK
2	54	100.0	342	2	GR9873
3	54	100.0	357	2	S21758
4	48	88.9	716	2	G82627
5	44	81.5	284	2	S25140
6	43	79.6	316	2	A45134
7	42	77.8	169	2	A12436
8	41	75.9	521	2	S62794
9	40	74.1	315	2	E71729
10	40	74.1	361	2	B56940
11	40	74.1	363	2	A56940
12	40	74.1	497	2	B97729
13	39	72.2	319	2	E87365
14	39	72.2	328	1	JN0453
15	39	72.2	364	2	AE3453
16	39	72.2	379	2	S23573
17	39	72.2	607	2	S01939
18	39	72.2	996	2	T50186
19	39	72.2	997	2	T39521
20	39	72.2	997	2	S63064
21	38	70.4	434	2	S23762
22	38	70.4	470	2	D75375
23	38	70.4	476	2	B82414
24	37	68.5	182	2	T49812
25	37	68.5	235	2	D89967
26	37	68.5	239	2	B89967
27	37	68.5	240	2	C89967
28	37	68.5	334	2	D82793
29	37	68.5	401	2	A48423

30	37	68.5	449	2	B71265	probable Mg2+ tran
31	37	68.5	462	2	H97292	UDP-N-acetyluramo
32	37	68.5	590	2	AB1411	autolysin, N-acety
33	37	68.5	817	2	T21336	hypothetical prote
34	37	68.5	830	2	T49270	receptor protein k
35	37	68.5	1254	2	T47141	hypothetical prote
36	37	68.5	1306	2	A70934	hypothetical glyci
37	37	68.5	1777	2	T34369	hypothetical prote
38	36	66.7	280	2	T28684	hypothetical prote
39	36	66.7	320	2	JC1311	cell protein precu
40	36	66.7	374	2	T09111	probable magnesium
41	36	66.7	381	2	JC7650	pectin lyase (EC 4
42	36	66.7	411	2	T15209	hypothetical prote
43	36	66.7	425	2	D64149	hypothetical prote
44	36	66.7	451	2	A82172	magnesium transpor
45	36	66.7	493	2	C97605	probable serine pr

#### ALIGNMENTS

##### RESULT 1

PRRSASK  
glutamy1 endopeptidase (EC 3.4.21.19) precursor - Staphylococcus aureus

N:Alternate names: staphylococcal serine proteinase  
C:Species: Staphylococcus aureus

C:Date: 04-Dec-1986 #sequence\_revision 30-Jun-1991 #text\_change 18-Jun-1999  
C:Accession: A26812; A00966

R:Carmona, C.; Gray, G.L.  
Nucleic Acids Res. 15, 6757, 1987

A:Title: Nucleotide sequence of the serine protease gene of Staphylococcus aureus, st  
A:Reference number: A26812; MUID:87316953; PMID:3306605

A:Accession: A26812

A:Molecule type: DNA

A:Residues: 1-336 <CAR>  
A:Cross-references: EMBL:Y00356; MID:946686; PIDN:CA68434.1; PID:946687

A:Experimental source: strain V8  
R:Drapeau, G.R.  
Can. J. Biochem. 56, 534-544, 1978

A:Reference number: A23824; MUID:78212487; PMID:96922

A:Accession: A00966

A:Molecule type: Protein

A:Residues: 69-108;110-124;126-144,'D',146-192,'T',194-228,'N',230-258,'Q',260,'D',26  
A:Experimental source: strain V8  
C:Comment: This extracellular proteolytic enzyme preferentially cleaves the peptide b  
C:Comment: This enzyme may be distantly related to the trypsin-type serine proteinase  
dues.

C:Superfamily: staphylococcal serine proteinase  
C:Keywords: hydrolase; serine proteinase

F:69-336/Product: staphylococcal serine proteinase #status experimental <ACT>  
F:119.161/Active site: His, Asp #status predicted  
F:237/Active site: Ser #status experimental

Query Match 100.0%; Score 54; DB 1; Length 336;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNNGSPVF 10  
DB 233 TGNNGSPVF 242

##### RESULT 2

hypothetical protein sspA [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: G89873

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mitutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: G89873  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-342 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13700850; PIDN:BA842146.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: ssrA  
 C:Superfamily: staphylococcal serine proteinase

Query Match 100.0%; Score 54; DB 2; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 0.13;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10  
 ||| ||||| |||  
 Db 233 TCGNSGSPVF 242

## RESULT 3

glutamic acid-specific endopeptidase - *Staphylococcus aureus*  
 C:Species: *Staphylococcus aureus*  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000  
 C:Accession: S21758  
 R:Yoshikawa, T.; Tsuruki, H.; Fujiwara, T.; Nakamura, E.; Yamoto, H.; Matsumoto, K.; Shi-  
 Biochim. Biophys. Acta 1121, 221-228, 1992  
 A:Title: Purification, characterization and gene cloning of a novel glutamic acid-specific  
 A:Reference number: S21758; MUID:92287954; PMID:1599945  
 A:Accession: S21758  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-357 <YOS>  
 A:Cross-references: GB:D00730; NID:g216970; PIDN:BA00630.1; PID:g216971  
 C:Superfamily: staphylococcal serine proteinase

Query Match 100.0%; Score 54; DB 2; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10  
 ||| ||||| |||  
 Db 233 TCGNSGSPVF 242

## RESULT 4

hypothetical protein XF1887 [Imported] - *Xylella fastidiosa* (strain 9a5c)  
 C:Species: *Xylella fastidiosa*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: G82627  
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen-  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82315; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: G82627  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-716 <STM>  
 A:Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AF84693.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Fachinani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigt  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, B  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1887

Query Match 88.9%; Score 48; DB 2; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 2.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9  
 ||| ||||| |||  
 Db 647 TCGNSGSPV 655

## RESULT 5

serine proteinase homolog - *Enterococcus faecalis*  
 C:Species: *Enterococcus faecalis*  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 17-Mar-2000  
 C:Accession: S25140  
 R:Su, Y.A.; Clewell, D.B.  
 submitted to the EMBL Data Library, June 1992  
 A:Description: A gene (SPIT) downstream of gele of *Enterococcus faecalis* OGI-10 resen  
 A:Reference number: S25140  
 A:Accession: S25140  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-284 <SU>  
 A:Cross-references: EMBL:Z12296; NID:g43337; PIDN:CA078168.1; PID:g43338  
 C:Superfamily: staphylococcal serine proteinase

Query Match 81.5%; Score 44; DB 2; Length 284;  
 Best Local Similarity 70.0%; Pred. No. 5.2;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10  
 ||| ||||| |||  
 Db 224 TCGNSGSPV 233

## RESULT 6

endopeptidase (EC 3.4.-.-), glutamate-specific - *Bacillus licheniformis*  
 C:Species: *Bacillus licheniformis*  
 C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Oct-1999  
 C:Accession: A45134; S23078  
 R:Kikudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S  
 J. Biol. Chem. 267, 23782-23788, 1992  
 A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s  
 A:Reference number: A45134; MUID:93054737; PMID:1429718  
 A:Accession: A45134  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-316 <KAK>  
 A:Cross-references: GB:D10060; NID:g216263; PIDN:BA00949.1; PID:d1001415; PID:g21626  
 A:Experimental source: ATCC 14580  
 A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBI:P:118785)  
 R:Swendsen, I.; Bredam, K.  
 Eur. J. Biochem. 204, 165-171, 1992

A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase  
 A:Reference number: S23078; MUID:92151519; PMID:1346764  
 A:Accession: S23078  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 95-316 <SVE>  
 C:Keywords: hydrolase

Query Match 79.6%; Score 43; DB 2; Length 316;  
 Best Local Similarity 88.9%; Pred. No. 8.5;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGNNGSPVF 10  
 || |||||  
 Db 258 GGNNGSPVF 266

## RESULT 7

hypothetical protein alr5049 [imported] - Nostoc sp. (strain PCC 7120)

A:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: A12436  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
 A:Reference number: AB1807; NID:21595285; PMID:11759840  
 A:Accession: A12436  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-169 <RUR>  
 A:Cross-references: GB:BA000019; PIDN:BA06748.1; PID:g17134187; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr5049

Query Match 77.8%; Score 42; DB 2; Length 169;  
 Best Local Similarity 88.9%; Pred. No. 6.7;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNNGSPV 9  
 |||||  
 Db 109 TGNNGSPV 117

## RESULT 8

probable lipoprotein D02\_orf521 - Mycoplasma pneumoniae (strain ATCC 29342)

N:Alternate names: MG395 homolog D02\_orf521  
 C:Species: Mycoplasma pneumoniae  
 A:Variety: ATCC 29342  
 C:Date: 27-Apr-1996 #sequence\_revision 01-Nov-1996 #text\_change 07-Dec-1999  
 C:Accession: S62794; S62798; S73576  
 R:Himmelfreid, R.; Hilbert, H.  
 Submitted to the EMBL Data Library, December 1995  
 A:Reference number: S62794

A:Accession: S62794  
 A:Molecule type: DNA  
 A:Residues: 1-521 <HIM>  
 A:Cross-references: EMBL:U043738; NID:g1209757; PIDN:AMC43664.1; PID:g1209771  
 R:Hilbert, H.; Himmelfreid, R.; Piagens, H.; Hermann, R.  
 Nucleic Acids Res. 24, 628-639, 1996  
 A:Title: Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae  
 A:Reference number: S62797; MUID:96177562; PMID:8604303  
 A:Accession: S62798

A:Molecule type: DNA  
 A:Residues: 1-50 <HIL>  
 A:Cross-references: EMBL:U043738  
 R:Himmelfreid, R.; Hilbert, H.; Piagens, H.; Pirk, E.; Li, B.C.; Hermann, R.  
 Nucleic Acids Res. 24, 4420-4449, 1996  
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae  
 A:Reference number: S73327; MUID:97105885; PMID:8948633  
 A:Accession: S73576

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-521 <HIL>  
 A:Cross-references: EMBL:AS000025; GB:U00089; NID:g1673918; PIDN:AA095898.1; PID:g167391  
 C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 19-Oct-1995  
 C:Accession: A56940  
 R:Kasai, K.; Takahashi, S.; Murakami, K.; Nakayama, K.  
 J. Biol. Chem. 270, 14471-14476, 1995  
 A:Title: Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse  
 A:Reference number: A56940; MUID:95301533; PMID:7540170  
 A:Accession: A56940  
 A:Superfamily: hypothetical protein MG068

Query Match 75.9%; Score 41; DB 2; Length 521;  
 Best Local Similarity 77.8%; Pred. No. 31;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGNNGSPVF 10  
 |||||  
 Db 423 GGNNGSPVF 431

## RESULT 9

protease DO (htra) RP186 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii  
 C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
 C:Accession: E71729  
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark  
 Nature 396, 133-140, 1998  
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
 A:Reference number: A71630; MUID:99039499; PMID:9823893  
 A:Accession: E71729  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-315 <AND>  
 A:Cross-references: GB:AJ235270; GB:AJ235269; NID:93860572; PIDN:CAA14652.1; PID:e134  
 A:Experimental source: strain Madrid E  
 C:Genetics:  
 A:Gene: htra; RP186

Query Match 74.1%; Score 40; DB 2; Length 315;  
 Best Local Similarity 88.9%; Pred. No. 27;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNNGSPV 9  
 |||||  
 Db 188 TGNNGSPV 196

## RESULT 10

integral membrane protein TGN38A - mouse

C:Species: Mus musculus (house mouse)  
 C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 19-Oct-1995  
 C:Accession: B56940  
 R:Kasai, K.; Takahashi, S.; Murakami, K.; Nakayama, K.  
 J. Biol. Chem. 270, 14471-14476, 1995  
 A:Title: Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse  
 A:Reference number: A56940; MUID:95301533; PMID:7540170  
 A:Accession: B56940  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-361 <KAS>  
 A:Cross-references: GB:D50032

Query Match 74.1%; Score 40; DB 2; Length 361;  
 Best Local Similarity 87.5%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNNGSPV 8  
 |||||  
 Db 138 TGNNGSPV 145

## RESULT 11

integral membrane protein TGN38B - mouse

C:Species: Mus musculus (house mouse)  
 C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 19-Oct-1995  
 C:Accession: A56940  
 R:Kasai, K.; Takahashi, S.; Murakami, K.; Nakayama, K.  
 J. Biol. Chem. 270, 14471-14476, 1995  
 A:Title: Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse  
 A:Reference number: A56940; MUID:95301533; PMID:7540170  
 A:Accession: A56940

A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-363 <KAS>  
A:Cross-references: GB:D50031

Query Match 74.1% Score 40; DB 2; Length 363;  
Best Local Similarity 87.5%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 8  
||| ||| |  
Db 140 TCGNSGSPV 147

## RESULT 12

B97729  
heat shock proteinase (EC 3.4.21.-) [imported] - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
C:Accession: B97729

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rd  
Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:21442074; PMID:11557893

A:Accession: B97729

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-497 <KUR>

A:Cross-references: GB:AE006914; PIDN:AL02772.1; PID:g15619287; GSPDB:GN00173

C:Genetics:

A:Gene: RC0234

C:Keywords: hydrolase; serine proteinase

Query Match 74.1% Score 40; DB 2; Length 497;  
Best Local Similarity 88.9%; Pred. No. 43;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9  
||| ||| |||  
Db 188 TCGASGSPV 196

## RESULT 13

E87365  
hypothetical protein CC0937 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: E87365

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwin, M.L.; Hart, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: E87365

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-319 <STO>

A:Cross-references: GB:AE005673; NID:g13422209; PIDN:AAK22921.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0937

Query Match 72.2% Score 39; DB 2; Length 319;  
Best Local Similarity 87.5%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GNSGSPV 10  
||| ||| |||  
Db 222 GNSGSPV 229

## RESULT 14

JN0453

dihydroorotate oxidase (EC 1.3.3.1) - g111 mushroom (Agrocycbe aegerita)  
N:Alternate names: dihydroorotate dehydrogenase  
C:Species: Agrocycbe aegerita  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: JN0453

R:Noel, T.; Labaree, J.  
Gene 122, 233-234, 1992

A:Title: Sequence of the URA1 gene encoding dihydroorotate dehydrogenase from the bas  
A:Reference number: JN0453; MUID:93083991; PMID:1452035

A:Accession: JN0453

A:Molecule type: DNA

A:Residues: 1-328 <NOE>

A:Cross-references: GB:M90295; NID:g166337; PIDN:AAA32636.1; PID:g166338

C:Genetics:

A:Gene: URA1

C:Superfamily: dihydroorotate oxidase

C:Keywords: flavoprotein; oxidoreductase; pyrimidine nucleotide biosynthesis

Query Match 72.2% Score 39; DB 1; Length 328;  
Best Local Similarity 70.0%; Pred. No. 41;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 10  
||| ||| |||  
Db 261 TGGLSGAPLF 270

RESULT 15  
AE3453  
dihydroorotate oxidase (EC 1.3.3.1) [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 19-Apr-2002  
C:Accession: AE3453

R:DelVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mufier, C.; Los, T.; Ivanov  
- Mazur, M.; Goltsman, E.; Selkov, P.H.; Hagius, S.; O'Callaghan, D.; Let  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
A:Reference number: AD3252; PMID:11736688

A:Accession: AE3453

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <KUR>

A:Cross-references: GB:AE008917; PIDN:AL52792.1; PID:g17983628; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEL1611

A:Map position: 1

C:Superfamily: dihydroorotate oxidase

C:Keywords: oxidoreductase

Query Match 72.2% Score 39; DB 2; Length 364;  
Best Local Similarity 70.0%; Pred. No. 46;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 10  
||| ||| |||  
Db 261 TGGLSGAPLF 270

Search completed: May 16, 2003, 13:48:09  
Job time: 2.73384 secs

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:47:08 ; Search time 0.909091 Seconds  
(without alignments)  
1060.959 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TGCNCGSPVF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications\_AA\*

1: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubppa/PCOT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubppa/PCOTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	10	9	US-10-008-355-25
2	54	100.0	52	9	US-10-008-355-3
3	54	100.0	52	9	US-10-008-355-4
4	54	100.0	699	9	US-10-008-355-8
5	54	100.0	712	9	US-10-008-355-2
6	54	100.0	732	9	US-10-008-355-6
7	48	88.9	9	9	US-10-008-355-26
8	48	88.9	716	9	US-10-008-355-7
9	48	88.9	720	9	US-10-008-355-9
10	44	81.5	734	9	US-10-008-355-5
11	37	68.5	74	10	US-09-864-761-48349
12	37	68.5	236	1	US-08-781-986A-5236
13	37	68.5	240	10	US-09-815-242-5351
14	37	68.5	254	10	US-09-815-242-12277
15	36	66.7	333	9	US-09-922-683-13
16	36	66.7	447	10	US-09-388-089B-2
17	36	66.7	475	10	US-09-388-089B-12
18	36	66.7	498	10	US-09-388-089B-11
19	36	66.7	1411	9	US-10-080-505-17

20	35	64.8	8	9	US-09-839-996-7	Sequence 7, Appl 1
21	35	64.8	8	9	US-09-839-996-8	Sequence 8, Appl 1
22	35	64.8	8	9	US-10-080-505-53	Sequence 53, Appl 1
23	35	64.8	8	9	US-10-080-505-54	Sequence 54, Appl 1
24	35	64.8	11	9	US-10-080-505-20	Sequence 20, Appl 1
25	35	64.8	499	10	US-09-864-761-35385	Sequence 35385, A
26	35	64.8	503	10	US-09-752-385-8	Sequence 8, Appl 1
27	35	64.8	550	10	US-09-905-657-2	Sequence 2, Appl 1
28	35	64.8	888	9	US-09-893-519A-73	Sequence 73, Appl 1
29	35	64.8	1391	9	US-10-080-505-11	Sequence 11, Appl 1
30	35	64.8	1391	9	US-10-080-505-15	Sequence 15, Appl 1
31	35	64.8	1394	9	US-09-839-996-13	Sequence 2, Appl 1
32	35	64.8	1394	9	US-10-080-505-2	Sequence 2, Appl 1
33	35	64.8	1395	9	US-10-080-505-7	Sequence 7, Appl 1
34	35	64.8	1434	9	US-10-080-505-9	Sequence 9, Appl 1
35	35	64.8	1436	9	US-10-080-505-13	Sequence 13, Appl 1
36	35	64.8	1541	9	US-09-839-996-3	Sequence 3, Appl 1
37	35	64.8	1541	9	US-10-080-505-3	Sequence 3, Appl 1
38	35	64.8	1545	9	US-09-839-996-4	Sequence 4, Appl 1
39	35	64.8	1545	9	US-10-080-505-4	Sequence 4, Appl 1
40	35	64.8	1702	9	US-09-839-996-5	Sequence 5, Appl 1
41	35	64.8	1702	9	US-10-080-505-5	Sequence 5, Appl 1
42	35	64.8	1702	9	US-09-839-996-6	Sequence 6, Appl 1
43	35	64.8	1848	9	US-10-080-505-6	Sequence 6, Appl 1
44	35	64.8	1848	9	US-10-091-504-976	Sequence 976, App
45	34	63.0	77	9		

## ALIGNMENTS

RESULT 1:  
US-10-008-355-25  
Sequence 25, Application US/10008355  
Patent No. US20020164759A1  
GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Potempa, Jan S  
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
FILE REFERENCE: 235.00440101  
CURRENT APPLICATION NUMBER: US/10/008.355  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 60/246,827  
PRIOR FILING DATE: 2000-11-08  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 25  
LENGTH: 10  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Consensus sequence  
US-10-008-355-25

Query Match 100.0%; Score 54; DB 9; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.0066;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCNCGSPVF 10  
Db 1 TGCNCGSPVF 10

RESULT 2  
US-10-008-355-3  
Sequence 3, Application US/10008355  
Patent No. US20020164759A1  
GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Potempa, Jan S  
APPLICANT: Banbula, Agnieszka  
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

FILE REFERENCE: 235.00440101  
CURRENT APPLICATION NUMBER: US/10/008,355  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 60/246,827  
PRIOR FILING DATE: 2000-11-08  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 3  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Porphyromonas gingivalis  
US-10-008-355-3

Query Match 100.0%; Score 54; DB 9; Length 52;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10  
DB 1 TCGNSGSPVF 10

RESULT 3  
US-10-008-355-4  
Sequence 4, Application US/10008355  
Patent No. US20020164759A1  
GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Potempa, Jan S  
APPLICANT: Banbula, Agnieszka  
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
FILE REFERENCE: 235.00440101  
CURRENT APPLICATION NUMBER: US/10/008,355  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 60/246,827  
PRIOR FILING DATE: 2000-11-08  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 4  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-10-008-355-4

Query Match 100.0%; Score 54; DB 9; Length 52;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10  
DB 1 TCGNSGSPVF 10

RESULT 4  
US-10-008-355-8  
Sequence 8, Application US/10008355  
Patent No. US20020164759A1  
GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Potempa, Jan S  
APPLICANT: Banbula, Agnieszka  
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
FILE REFERENCE: 235.00440101  
CURRENT APPLICATION NUMBER: US/10/008,355  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 60/246,827  
PRIOR FILING DATE: 2000-11-08  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 8  
LENGTH: 699  
TYPE: PRT  
ORGANISM: Porphyromonas gingivalis

US-10-008-355-8

Query Match 100.0%; Score 54; DB 9; Length 699;  
Best Local Similarity 100.0%; Pred. No. 0.51;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10  
DB 644 TCGNSGSPVF 653

RESULT 5  
US-10-008-355-2  
Sequence 2, Application US/10008355  
Patent No. US20020164759A1  
GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Potempa, Jan S  
APPLICANT: Banbula, Agnieszka  
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
FILE REFERENCE: 235.00440101  
CURRENT APPLICATION NUMBER: US/10/008,355  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 60/246,827  
PRIOR FILING DATE: 2000-11-08  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 712  
TYPE: PRT  
ORGANISM: Porphyromonas gingivalis  
US-10-008-355-2

Query Match 100.0%; Score 54; DB 9; Length 712;  
Best Local Similarity 100.0%; Pred. No. 0.52;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10  
DB 644 TCGNSGSPVF 653

RESULT 6  
US-10-008-355-6  
Sequence 6, Application US/10008355  
Patent No. US20020164759A1  
GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Potempa, Jan S  
APPLICANT: Banbula, Agnieszka  
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
FILE REFERENCE: 235.00440101  
CURRENT APPLICATION NUMBER: US/10/008,355  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 60/246,827  
PRIOR FILING DATE: 2000-11-08  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 6  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Shewanella putrefaciens  
US-10-008-355-6

Query Match 100.0%; Score 54; DB 9; Length 732;  
Best Local Similarity 100.0%; Pred. No. 0.54;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10  
DB 662 TCGNSGSPVF 671



RESULT 7  
US-10-008-355-26  
; Sequence 26, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 26  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus sequence  
US-10-008-355-26

Query Match 88.9%; Score 48; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.3e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9  
Db 1 TCGNSGSPV 9

RESULT 8  
US-10-008-355-7  
; Sequence 7, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 716  
; TYPE: PRT  
; ORGANISM: *Xylella fastidiosa*  
US-10-008-355-7

Query Match 88.9%; Score 48; DB 9; Length 716;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9  
Db 647 TCGNSGSPV 655

RESULT 9  
US-10-008-355-9  
; Sequence 9, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: *Porphyromonas gingivalis*  
US-10-008-355-9

Query Match 88.9%; Score 48; DB 9; Length 720;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9  
Db 651 TCGNSGSPV 659

RESULT 10  
US-10-008-355-5  
; Sequence 5, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 734  
; TYPE: PRT  
; ORGANISM: *Shewanella putrefaciens*  
US-10-008-355-5

Query Match 81.5%; Score 44; DB 9; Length 734;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSP 8  
Db 665 TCGNSGSP 672

RESULT 11  
US-09-864-761-48349  
; Sequence 48349, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 48349  
LENGTH: 74  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL11716.16  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9  
OTHER INFORMATION: EST\_HUMAN HIT: BE002805.1, EVALUATE 8.00e-39  
OTHER INFORMATION: SWISSPROT HIT: Q9ZKD2, EVALUATE 3.60e+00  
US-09-664-761-48349

Query Match 68.5%; Score 37; DB 10; Length 74;  
Best Local Similarity 75.0%; Pred. No. 30;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 GNGSPV 9  
II:IIII:  
Db 52 GGHSSPT 59

RESULT 12  
US-08-781-986A-5236  
Sequence 5236, Application US/08781986A  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,986A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PB248PP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8512  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 5236:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 236 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-781-986A-5236

Query Match 68.5%; Score 37; DB 1; Length 236;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNGSPV 9  
IIIIII:  
Db 188 GNGSPV 194

RESULT 13  
US-09-815-242-5351  
Sequence 5351, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5351  
LENGTH: 240  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5351

Query Match 68.5%; Score 37; DB 10; Length 240;  
Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 GNSGSPV 9  
|111111|  
Db 191 GNSGSPV 197

RESULT 14  
US-09-815-242-12277  
; Sequence 12277, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes In  
; FILE REFERENCE: ELTRA, 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12277  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12277

Query Match 68.5%; Score 37; DB 10; Length 254;  
Best Local Similarity 100.0%; Pred. No. 1,1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNSGSPV 9  
|111111|  
Db 205 GNSGSPV 211

RESULT 15  
US-09-922-683-13  
; Sequence 13, Application US/09922683  
; Publication No. US20020192793A1  
; GENERAL INFORMATION:  
; APPLICANT: DECKER, Heinrich  
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
; PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUDESCENS  
; GLA O AND THEIR USE

NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/922,683  
; FILING DATE: 07-Aug-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/194,905  
; FILING DATE: 1999-12-01  
; APPLICATION NUMBER: DE 19622783.6  
; FILING DATE: 07-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granados, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 026083/0193  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 393 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: <Unknown>  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
US-09-922-683-13

Query Match 66.7%; Score 36; DB 9; Length 393;  
Best Local Similarity 85.7%; Pred. No. 2,4e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GNSGSP 8  
|111111|  
Db 91 GNSGAP 97

Search completed: May 16, 2003, 14:03:03  
Job time: 1.90909 secs



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:33:37 ; Search time 1.46769 Seconds  
(without alignments)  
1403:888 Million cell updates/sec

Title: US-10-008-355-25  
Perfect score: 54  
Sequence: 1 TCGNSGSPVF 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	54	100.0	316	2	09FBG1	09fbg1 staphylococ
2	54	100.0	342	16	099V45	099v45 staphylococ
3	54	100.0	357	2	004186	004186 staphylococ
4	49	90.7	217	2	09AJX0	09ajx0 staphylococ
5	48	88.9	716	16	09PC94	09pc94 xyliella fas
6	44	81.5	284	2	047809	047809 enterococcu
7	44	81.5	289	16	09EM08	09em08 rhizobium l
8	44	81.5	398	11	08YE75	08ye75 mus musculu
9	44	81.5	613	11	09D219	09d219 mus musculu
10	43	79.6	1008	16	09BK09	09bk09 rhizobium l
11	42	77.8	169	16	08YM87	08ym87 anabaena sp
12	42	77.8	458	16	08XHM4	08xhm4 clostridium
13	40	74.1	315	16	09ZDX8	09zdx8 rickettsia
14	40	74.1	353	11	06Z313	06z313 mus musculu
15	40	74.1	363	11	06Z314	06z314 mus musculu
16	40	74.1	497	16	09ZJ35	09zj35 rickettsia

17	40	74.1	639	16	09ZLB9	09zlb9 rhizobium m
18	39	72.2	319	16	09A9N9	09a9n9 caulobacter
19	39	72.2	364	16	08YFB1	08yfb1 bruceella me
20	39	72.2	379	3	09Y891	09y891 colletotric
21	39	72.2	396	12	08YAO4	08yao4 apple stem
22	39	72.2	499	10	09LKT0	09lkt0 arabidopsis
23	39	72.2	996	3	09P7S1	09p7s1 schizosacch
24	39	72.2	997	3	074325	074325 schizosacch
25	38	70.4	253	2	P96151	P96151 vibrio chol
26	38	70.4	359	16	09AB89	09ab89 rhizobium l
27	38	70.4	394	12	08OR15	08or15 apple stem
28	38	70.4	440	17	08TZS8	08tzs8 pyrococcus
29	38	70.4	441	10	09FMY1	09fmy1 arabidopsis
30	38	70.4	470	16	09RTZ8	09rtz8 deinococcus
31	38	70.4	476	16	09KLD4	09kld4 vibrio chol
32	38	70.4	568	12	089504	089504 cocksfoot m
33	38	70.4	568	12	066149	066149 cocksfoot m
34	38	70.4	568	12	09E960	09e960 cocksfoot m
35	38	70.4	579	16	08RE47	08re47 fusobacteri
36	38	70.4	593	10	09FYL4	09fyl4 arabidopsis
37	38	70.4	627	10	09SRP2	09srp2 arabidopsis
38	38	70.4	1097	10	08RY22	08ry22 arabidopsis
39	38	70.4	1335	2	09LA58	09la58 escherichia
40	38	70.4	1335	2	09LA54	09la54 escherichia
41	37	68.5	235	2	09KH51	09kh51 staphylococ
42	37	68.5	235	2	09EPD8	09epd8 staphylococ
43	37	68.5	235	16	099T60	099t60 staphylococ
44	37	68.5	239	2	09KH49	09kh49 staphylococ
45	37	68.5	239	16	053782	053782 staphylococ

## ALIGNMENTS

## RESULT 1

09FBG1 ID 09FBG1 PRELIMINARY; PRT; 316 AA.  
AC 09FBG1;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE GlutamyI endopeptidase.  
GN PROM.  
OS Staphylococcus warneri.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Staphylococcus.  
OX NCBI\_TaxID=1292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kikikawa M.;  
RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL: A793885; CAC06168.1; -  
DR MEROPS; S01.269; -  
DR InterPro: IPR001254; Ser\_protease\_Try.  
DR InterPro: IPR00126; Ser\_protease\_V8.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00839; VBPROTEASE.  
DR SMART: SM00020; TRYP\_SPC; 1.  
DR PROSITE: PSS0240; TRYP\_SIN\_DOM; 1.  
DR PROSITE: PS00673; V8\_SER; 1.  
KW Hydrolase; Serine protease.  
SQ SEQUENCE 316 AA; 4E997A5A11IDB40 CRC64;

Query Match 100.0%; Score 54; DB 2; Length 316;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10  
 |||||||||  
 DB 231 TCGNSGSPVF 240

## RESULT 2

099V45 PRELIMINARY: PRT: 342 AA.  
 AC 099V45:  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE Serine protease, V8 protease, glutamyl endopeptidase.  
 GN SSPA OR SAV1048 OR SA0901.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OC Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 NCBI\_TaxID=158878, 158879;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECTES-S aureus (strain Mu50), and S. aureus (strain N315);  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,  
 RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshoyama A.,  
 RA Mizutani-U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekinaka M., Hirakawa H., Kubara S., Goto S., Yabuzaki J.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hirumatsu K.,  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus."  
 RL Lancel 357:1225-1240(2001).  
 CC -:- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
 DR EMBL: AP003361; BAB57210.1; -;  
 DR EMBL: AP003132; BAB42146.1; -;  
 DR MEROPS: S01.269; -;  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR00126; Ser\_protease\_V8.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00839; V8PROTEASE.  
 DR SMART: SM00020; TRYPSIN\_SPEC.1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00672; V8\_HIS.1.  
 DR PROSITE: PS00673; V8\_SER.1.  
 KW Hydrolase: Protease: Serine protease: Complete proteome.  
 SQ SEQUENCE 342 AA: 36977 MW: 5AEP42DCE01C4B24 CRC64;  
 Query Match 100.0%; Score 54; DB 16; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 0.17;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TCGNSGSPVF 10  
 |||||||||  
 DB 233 TCGNSGSPVF 242

## RESULT 3

004186 PRELIMINARY: PRT: 357 AA.  
 AC 004186:  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Glutamic acid specific protease prepeptide (EC 3.4.21.19).  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yoshikawa K., Suzuki H., Fujiwara T., Nakamura E., Iwamoto H.,  
 RA Matsumoto K., Shin M., Yoshida N., Teraoka H.;

RT "Purification, characterization and gene cloning of a novel glutamic  
 RT acid-specific endopeptidase from Staphylococcus aureus atcc 12600.";  
 RL Biochim. Biophys. Acta 1121:221-228(1991).  
 DR EMBL: D00730; BAA00630.1; -;  
 DR MEROPS: S01.269; -;  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR000126; Ser\_protease\_V8.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00839; V8PROTEASE.  
 DR PROSITE: PS50240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00672; V8\_HIS.1.  
 DR PROSITE: PS00673; V8\_SER.1.  
 KW Hydrolase: Protease: Serine protease.  
 FT CHAIN 69 357  
 SQ SEQUENCE 357 AA: 38651 MW: 58AA94AE371E2577 CRC64;

Query Match 100.0%; Score 54; DB 2; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 0.17;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10  
 |||||||||  
 DB 233 TCGNSGSPVF 242

## RESULT 4

09AJX0 PRELIMINARY: PRT: 217 AA.  
 AC 09AJX0:  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE Extracellular serine proteinase precursor (Fragment).  
 GN ESP.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6746;  
 RA Dudin G.;  
 RT "Staphylococcus epidermidis extracellular serine proteinase."  
 RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AJ305145; CAC27157.1; -;  
 DR HSSP: P09331; IEXF.  
 DR MEROPS: S01.269; -;  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR00126; Ser\_protease\_V8.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00839; V8PROTEASE.  
 DR PROSITE: PS50240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00673; V8\_SER.1.  
 KW Hydrolase: Serine protease: Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL 1 1  
 FT CHAIN 2 217  
 FT CHAIN 2 217  
 SQ SEQUENCE 217 AA: 23667 MW: FB98886D453B8B7 CRC64;  
 Query Match 90.7%; Score 49; DB 2; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 0.79;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGNSSPVF 10  
 |||||||||  
 DB 167 GGNSSPVF 175

## RESULT 5

09PC94 PRELIMINARY: PRT: 716 AA.  
 AC 09PC94:  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)

```

DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
GN Xf1887.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RA MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnsels J.D., Junqueira M.L., Kemper E.L., Kitejima J.P.,
RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.V., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Pexoco B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quagga R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA da Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Sawasaki H.E.,
RA de Souza A.P., Terezi M.F., Trufi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Secubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AF004008; AAF84653.1; -
DR MEROPS: S46.001; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BDFC CRC64;

Query Match 88.9%; Score 48; DB 16; Length 716;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
Db 647 TGNSSGSPV 655

RESULT 6
Q47809 PRELIMINARY; PRT; 284 AA.
AC Q47809;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Staphylococcal serine proteinase homologue.
GN SPRE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Enterococcaceae; Enterococcus.
NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OGI-10;
RA Su Y.A., Clewell D.B.;
RT "A gene (sprt) downstream of gele of Enterococcus faecalis OGI-10
resembles serine proteinase determinant of Staphylococcus aureus

```

```

RT strain V8.";
RL submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z12296; CAA78168.1; -
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR00126; Ser_proteas_V8.
DR PRINTS: PR00839; VBPROTEASE.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
DR PROSITE: PS00672; V8_HIS; 1.
DR PROSITE: PS00673; V8_SER; UNKNOWN_1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 284 AA; 31063 MW; DAF0312BEE778415 CRC64;

Query Match 81.5%; Score 44; DB 2; Length 284;
Best Local Similarity 70.0%; Pred. No. 7.8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 10
Db 224 TGNSSGSPV 233

RESULT 7
Q98M08 PRELIMINARY; PRT; 289 AA.
AC Q98M08;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Hypothetical protein ms18587.
GN MS18587.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RA MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002995; BAB48055.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 289 AA; 31310 MW; 8947CDDFC6BC0253 CRC64;

Query Match 81.5%; Score 44; DB 16; Length 289;
Best Local Similarity 88.9%; Pred. No. 7.9;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGNSSGSPV 10
Db 194 GGNSSGSPV 202

RESULT 8
Q8VE75 PRELIMINARY; PRT; 398 AA.
AC Q8VE75;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Similar to RIKEN cDNA 4632417K18 gene (fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC019638; AAH19638.1;  
 FT NON TER  
 SO SEQUENCE 398 AA: 45268 MW: 51FED8CE693AC1B7 CRC64;

Query Match  
 Best Local Similarity 81.5%; Score 44; DB 11; Length 398;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GCGSGSPVF 10  
 11:11111111  
 Db 325 GCGSGSPVF 333

## RESULT 9

O992L9 PRELIMINARY: PRT: 613 AA.

ID O9D2L9  
 AC O9D2L9  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE 4632417K18RIK protein.  
 GN 4632417K18RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=SKIN;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinaigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kaota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaudo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlski S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK019499; BAB31763.1;  
 DR MGP: MGI:1915508; 4632417K18RIK.  
 SO SEQUENCE 613 AA: 69948 MW: F926114F705A6399 CRC64;

Query Match  
 Best Local Similarity 81.5%; Score 44; DB 11; Length 613;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GCGSGSPVF 10  
 11:11111111  
 Db 540 GCGSGSPVF 548

## RESULT 10

O98KC9 PRELIMINARY: PRT: 1008 AA.

ID O98KC9  
 AC O98KC9  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein ml1534.

GN ML1534.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP002997; BAB48885.1;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 1008 AA: 100023 MW: AC3F287D0B94C53 CRC64;

Query Match  
 Best Local Similarity 79.6%; Score 43; DB 16; Length 1008;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10  
 1111111111  
 Db 442 TCGNSGSPVF 451

## RESULT 11

O8YM87 PRELIMINARY: PRT: 169 AA.

ID O8YM87  
 AC O8YM87  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein Alr5049.  
 GN ALR5049.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21595285; PubMed=11759840;  
 RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iritaguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsuno M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL: AP003598; BAB76748.1;  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 169 AA: 17427 MW: 0A610CC04EAC48A CRC64;

Query Match  
 Best Local Similarity 77.8%; Score 42; DB 16; Length 169;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 9  
 1111111111  
 Db 109 TCGNSASPV 117

## RESULT 12

O8XHM4 PRELIMINARY: PRT: 458 AA.

ID O8XHM4  
 AC O8XHM4  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)



DE UDP-N-acetylmuramoylalanine D-glutamate ligase.  
 GN MORD OR CPE2459.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 CC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / TYPE A;  
 RX PubMed=11792842;  
 RA Shimizu T., Ohnani K., Hirakawa H., Ohshima K., Yamashita A.,  
 Shiba T., Ogasawara N., Hattori M., Kunara S., Hayashi H.,  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 flesh-eater."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL: AP003194; BAB82165.1; -.  
 DR InterPro: IPR000713; Mur\_Ligase.  
 DR InterPro: IPR004101; Mur\_Ligase\_C.  
 DR Pfam: PF01225; Mur\_Ligase\_1.  
 DR Pfam: PF02875; Mur\_Ligase\_C; 1.  
 DR TIGRFAMs: TIGR01087; murD; 1.  
 KM Ligase; Complete proteome.  
 SQ SEQUENCE 458 AA; 51513 MW; CDB19A3365EEAE64 CRC64;

Query Match 77.8%; Score 42; DB 16; Length 458;  
 Best Local Similarity 70.0%; Pred. No. 28;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGNSSGSPV 10  
 DB 148 TGNIGTLPF 157

## RESULT 13

09ZDX8 PRELIMINARY; PRT; 315 AA.  
 AC 09ZDX8;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Protease DO (HTRA).  
 GN Rpl86.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=782;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MADRID E;  
 RC MEDLINE=99039499; PubMed=9823893;  
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sicheit-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria."  
 RL Nature 396:133-140(1998).  
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
 DR EMBL: AJ235270; CAA14652.1; -.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001940; Protease2C.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PRO0834; PROTEASES2C.  
 DR PROSITE: PS50106; PDZ; 1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
 KM Hydrolyase; Serine protease; Complete proteome.  
 SQ SEQUENCE 315 AA; 35033 MW; 2D91A0D54FEB9A1 CRC64;

Query Match 74.1%; Score 40; DB 16; Length 315;  
 Best Local Similarity 88.9%; Pred. No. 44;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGNSSGSPV 9

DB 188 TGNSSGSPV 196

## RESULT 14

06Z313 PRELIMINARY; PRT; 353 AA.  
 AC 06Z313;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Trans-golgi network integral membrane protein TGN38A precursor (Trans-  
 golgi network protein 1) (TGN38 homolog).  
 GN TGN1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR; TISSUE=BRAIN;  
 RX MEDLINE=95301533; PubMed=7540170;  
 RA Kasai K., Takahashi S., Murakami K., Nakayama K.;  
 RT "Strain-specific presence of two TGN38 isoforms and absence of TGN41  
 RT in mouse."  
 RL J. Biol. Chem. 270:14471-14476(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Strauberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND  
 CC FROM TRANS-GOLGI NETWORK.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-  
 CC GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL  
 CC SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.  
 CC EMBL: D50031; BAA08757.1; -.  
 DR EMBL: BC009143; AAH09143.1; -.  
 DR MGD: NGI:105080; Tgn1.  
 KM Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.  
 FT SIGNAL 1  
 FT CHAIN 18 353  
 FT FT 18 353  
 FT FT 18 353  
 FT DOMAIN 18 298  
 FT TRANSMEM 299 319  
 FT DOMAIN 320 353  
 FT DOMAIN 346 349  
 FT DOMAIN 131 178  
 FT REPEAT 131 138  
 FT REPEAT 139 146  
 FT REPEAT 147 154  
 FT REPEAT 155 162  
 FT REPEAT 163 170  
 FT REPEAT 171 178  
 FT CARBOHYD 110 110  
 FT CARBOHYD 293 293  
 SO SEQUENCE 353 AA; 37848 MW; 95C340C2FAA1BE3 CRC64;

Query Match 74.1%; Score 40; DB 11; Length 353;  
 Best Local Similarity 87.5%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGNSSGSPV 8  
 DB 138 TGNSSGSPV 145

RESULT 15  
 ID 06Z314 PRELIMINARY; PRT; 363 AA.  
 AC 06Z314;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMURel. 01, last sequence update)  
 DT 01-DEC-2001 (TREMURel. 19, last annotation update)  
 DE Trans-golgi network integral membrane protein TGN38B precursor (Trans-  
 golgi network protein 2) (TGN38 homolog).  
 GN TGN38 OR TGN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ICR: TISSUE-BRAIN;  
 RX MEDLINE=95301533; PubMed=7540170;  
 RA Kasai K., Takahashi S., Murakami K., Nakayama K.;  
 RT "Strain-specific presence of two TGN38 isoforms and absence of TGN41  
 in mouse."  
 RL J. Biol. Chem. 270:14471-14476(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND  
 FROM TRANS-GOLGI NETWORK.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-  
 GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL  
 SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.  
 CC -1- MISCELLANEOUS: NOT FOUND IN STRAINS BALB/C, C57BL/6 AND DBA/2.  
 DR EMBL: D50032; BAA08758.1; -;  
 DR MGD: MGI:105079; Tgln2.  
 KW Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.  
 FT SIGNAL 1 17 POTENTIAL.  
 FT CHAIN 18 363 TRANS-GOLGI NETWORK INTEGRAL MEMBRANE  
 FT DOMAIN 18 308 PROTEIN TGN38B.  
 FT TRANSMEM 309 329 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 330 363 POTENTIAL.  
 FT SITE 356 359 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 133 188 ENDOCYTOSIS SIGNAL (BY SIMILARITY).  
 FT REPEAT 133 140 7 X 8 AA TANDEM REPEATS.  
 FT REPEAT 141 148 1.  
 FT REPEAT 149 156 2.  
 FT REPEAT 147 164 3.  
 FT REPEAT 165 172 4.  
 FT REPEAT 173 180 5.  
 FT REPEAT 181 188 6.  
 FT CARBOHYD 303 303 7.  
 SQ SEQUENCE 363 AA; 38621 MW; 2826FA9E958C5C27 CRC64; N-LINKED (GLCNAC... ) (POTENTIAL).

Query Match 74.1%; Score 40; DB 11; Length 363;  
 Best Local Similarity 87.5%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGN38SP 8  
 |||||  
 Db 140 TGN38SP 147

Search completed: May 16, 2003, 13:47:01  
 Job time : 4.46769 secs

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 07:32:23 ; Search time 84.8184 Seconds

(without alignments)  
5071.208 Million cell updates/sec

Title: US-10-008-355-2\_COPY\_522\_712

Perfect score: 1003

Sequence: 1 SKSVIAAARAIQADAMANAY.....LFMDKWCQCPRLLQELKLI 191

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Dgapop 6.0 , Dgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```

-MODEL=frame+.p2n.model -DEV=xjh
-O=/cgnr2_1/USPFO.spool/US10008355/runat_16052003_110359_2525/app_query.fasta_1.1230
-DB=N_Geneseq_101002 -QEXT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blsum2 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000
-USER=US10008355 @CGN.1.1.275 @runat_16052003_110359_2525 -NCPU=6 -ICPU=3
-NO_XLPPY -NO_MMAP -LARGEOUTRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database :

```

1: N_Geneseq_101002:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT:*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1003	100.0	2139	AA143635	Porphyromonas ging
2	90	9.0	2136	AA111244	Neisseria meningit
3	89.5	8.9	2794	AAH17774	Human cDNA sequenc
4	89.5	8.9	3259	AA159418	Human polynucleoti
5	89.5	8.9	3514	AAH16079	Human cDNA sequenc
6	89.5	8.9	3899	AAAS6370	Human encoding novel
7	89.5	8.9	3915	AA161204	Human polynucleoti
8	89.5	8.9	6258	AAAS1791	DNA encoding novel
9	89.5	8.9	10862	AAV34396	Yellow fever infec
10	89.5	8.9	10892	AAE83821	Chimeric DNA from
11	86	8.6	2365389	ABA90521	Genomic sequence o
12	85.5	8.5	651	ABO65948	Arabidopsis thalia
13	85.5	8.5	972	AAAC6293	Arabidopsis thalia
14	83	8.3	606	AAH53441	S. epidermidis ope
15	83	8.3	731	AAE91391	N. meningitidis (s
16	83	8.3	731	ABK37771	DNA sequence upstr
17	83	8.3	1083	ABN92057	Staphylococcus epi
18	83	8.3	3003	AAH54900	S. epidermidis gen
19	82.5	8.2	319630	ABO67194	Listeria innocua c
20	82	8.2	3253	AAAC75951	Human ORFX ORF1506
21	81.5	8.1	720	AAAS1714	Staphylococcus aur
22	81.5	8.1	765	AAAS5452	Staphylococcus aur
23	81.5	8.1	927	ABN66304	Streptococcus poly
24	81.5	8.1	2155561	ABN71527	Streptococcus poly
25	81	8.1	849	AAAD01293	Escherichia canis im
26	81	8.1	849	ABK68853	Escherichia canis im
27	81	8.1	2037	AAAD01295	DNA encoding novel
28	80.5	8.0	3951	AAAS82910	Drosophila melanog
29	80.5	8.0	14105	ABL29585	Drosophila melanog
30	80.5	8.0	17769	ABL29584	Staphylococcus aur
31	80	8.0	1019	AAAL5963	Staphylococcus aur
32	80	8.0	1558	AAO27988	Protease from S. A
33	80	8.0	1585	AAO27987	Protease from S. A
34	79.5	7.9	861	ABK75469	Bacillus lichenifo
35	79.5	7.9	2244	ABK75469	Nucleic acid seque
36	79.5	7.9	2244	ABK75469	Genomic sequence o
37	79	7.9	1833	AAE90521	Konjak mosaic viru
38	78.5	7.8	1497	AAZ54262	Neisseria meningit
39	78.5	7.8	6444	AAV74544	Staphylococcus aur
40	78	7.8	2017	AAH99629	Human protein enco
41	78	7.8	2541	AAAC42943	Arabidopsis thalia
42	78	7.8	2643	AAK94354	Human full-length
43	78	7.8	3192	AAH16193	Human cDNA sequenc
44	78	7.8	10976	ABL50890	Japanese encephali
45	77.5	7.7	905	AAAS0158	DNA encoding novel

## ALIGNMENTS

RESULT 1	
AA143635	
AA143635 standard; DNA; 2139 BP.	
AA143635:	
05-SEP-2002 (first entry)	
XX	Porphyromonas gingivallis dipeptidylpeptidase-7 (DPP-7) coding sequence.
XX	Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;
KW	DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW	periodontitis.
XX	
OS	Porphyromonas gingivallis.
XX	
Key	Location/Qualifiers
FT	CDS 1..2139

FT		/tag= a	
XX		/product= "Porphyromonas gingivalis DPP-7"	
XX			
PN		W0200238742 .n2.	
PD		16-MAY-2002.	
PF		08-NOV-2001; 2001WO-US46782.	
PR		08-NOV-2000; 2000US-246827P.	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.		
PI	Travis J, Potempa JS, Banbula A, Bugno M;		
PX	WPI: 2002-490075/52.		
DR	P-PSDB; AA015205.		
XX			
PT	Novel isolated dipeptidylpeptidase useful for identifying inhibitor of		
PT	the dipeptidylpeptidase for protecting an animal from periodontal		
PT	disease caused by Porphyromonas gingivalis -		
XX	Claim 11; Fig 4; 65pp; English.		
PS			
XX	The invention comprises the amino acid and coding sequence of the		
CC	Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7		
CC	enzyme has amidolytic activity for cleavage of a peptide bond between the		
CC	second and third amino acids from the N-terminal end of a target peptide.		
CC	The DPP-7 target peptide has an aliphatic or aromatic residue as a		
CC	substituent on the alpha-carbon atom of the second amino acid from the		
CC	N-terminal end. The DPP-1 protein and DNA sequences of the invention are		
CC	useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for		
CC	reducing the growth of a bacterium and protecting an animal from a		
CC	periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis		
CC	or periodontitis). The present DNA sequence encodes the Porphyromonas		
CC	gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.		
SQ	Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other:		
	Alignment Scores:		
	Pred. No.: 2 04e-113 Length: 2139		
	Score: 1003.00 Matches: 191		
	Percent Similarity: 100.00% Conservative: 0		
	Best Local Similarity: 100.00% Mismatches: 0		
	Query Match: 100.00% Indels: 0		
	DB: 24 Gaps: 0		
	US-10-008-335-2_COPY_522_712 (1-191) x AAL43635 (1-2139)		
OY	1 SerIysSerValIIleAlaIalaArGAlaIIleGlNaIaaSPAlaMeTAlaAsnaIATyr 20		
Db	1564 TCcAAAGGGGTATTACTGCTGCTCGGCGTATTACAGCCGCATGGCAGTCGAATGCCAT 1623		
OY	21 AlaIIleGIuLysGIlySArGLeuPhePhaIaGIlyLeuAArgIuMetTYrPrOGIlyArg 40		
Db	1624 GCCATTGAGAAGGCAAGCCCTCTTTCTTTGCCGTTGGCTGAGATCTACCCEGGAGCT 1683		
OY	41 AlaLeuProSerSerSpIlaAsnPhetHmeTArMetSerIyrgIySerIIelYSglTYr 60		
Db	1684 GCCTCTCCGAGCGCATGCACTTCACCATCGTAGTGAGCTTAAGGCTCATCAAGGATAT 1743		
OY	61 GIuPrOGIlaSPclYlaATPrTYrASnTYrhISThrThrGIlySGIlyVaIIleuGIuLYs 80		
Db	1744 GAACCGCAGAGCGGTGCTGTACAACATCATATCGACAGGCAAGGGGCGATTGAGAGAG 1803		
OY	81 GlNaSPProLySSerASPgluPheAlaValGIingluAsnIIleLeuaspLeuPheaRgThr 100		
Db	1804 CAAGATCTCAAGAGCGCATGTAGTTGGCTGCATCAGAGATAATCTCCGACCTTCCGCACC 1863		
OY	101 LysASnTYrGIlyARgTYrAlaGIuASngIyGIleuHisTIleAlaPheLeuSerASnSn 120		
Db	1864 AAAAACtATGTGTGCTATGCGAGAACGGTCACTTCATATCGCTTTCTTCTATCGAACACAC 1923		

QY	121	AspIlethrhgLygIyaAnSerGlySerProValPheAspLyAsnGlyVArgheuIleGly	140
Db	1924	GACATTCACGGGGCGGTAACTCCGGTAGCCCGTATTCGATPAGACGCGCTGTGATCCGT	1983
QY	141	LeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeu	160
Db	1984	CTTGCTTTTCATGGCACCCTGGAAAGCATATGAGTGGACATCGAGTTCGAACCCGATCTG	2043
QY	161	GlnArgThrIleSerValAspIleArgTyValLeuPheMetIleAspLysTrpGlyGln	180
Db	2044	CAGCCACAAATCACCAGCTGGAGCATCCGCTACGTTCTCTCATGATGTGCAAAATGGGGTCAG	2103
QY	181	CysProArgLeuIleGlnGluLeuLysLeuIle	191
Db	2104	TGCCCCGCTCATCCCAAGAGCTGAAGTTGATC	2136
RESULT 2			
AAFI1244			
ID	AAFI1244	standard; DNA: 2136 BP.	
XX	AAFI1244:		
XX	02-SEP-1996	(first entry)	
XX			
XX	Neisseria meningitidis strain BZ163	transferrin receptor Tbp2 gene.	
XX			
XX	Transferin receptor: Tbp2 subunit; deletion mutant; vaccine:		
KM	passive immunisation; immunotherapy: IM2169; IM2394; ss.		
XX			
OS	Neisseria meningitidis (strain BZ163).		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	1..2136	
FT		/*tag= a	
FT	sig_peptide	1..60	
FT		/*tag= b	
FT	mat_peptide	61..2133	
FT		/*tag= c	
XX			
XX	W09533049-A2.		
XX			
PD	07-DEC-1995:		
XX			
PF	30-MAY-1995:	95WO-FR00701.	
XX			
PR	31-MAY-1994:	94FR-0006594.	
XX			
PA	(INMR ) PASTEUR MERIEUX SERUMS & VACCINS.		
XX	(TRGE ) TRANSGENE SA.		
PI	Jacobs E, Legrain M, Mazarin V, Lissolo L;		
PI	Millet MBJ;		
XX			
DR	WPI: 1996-030562/03.		
XX	P-PSDB: AAR8649.		
XX			
PT	Polypeptide(s) for vaccination against Neisseria meningitidis group		
PT	B - comprising deletion mutants of transferrin receptor Tbp2		
PT	subunit		
XX			
XX	Disclosure: Page 82-87; 114pp; French.		
XX			
CC	The present sequence is that of the N.meningitidis strain BZ163		
CC	transferrin Tbp2 subunit gene. The Tbp2 polypeptide has three		
CC	domains: an N-terminal domain, a hinge domain and a C-terminal		
CC	domain; deletion mutants in which at least one of the domains is		
CC	partially or totally deleted are claimed, provided that the first		
CC	and second domains are not simultaneously partially or totally		
CC	deleted. The positions of the 3 domains in BZ163 are defined by		
CC	alignment with the IM2169 sequence. The deletion mutant polypeptides		
CC	of the invention can generate an immune response against		
CC	N.meningitidis.		
XX			

50	Sequence 2136 BP; 695 A; 474 C; 523 G; 444 T; 0 other.
Alignment Scores:	
Pred. No.:	0.732
Score:	90.00
Percent Similarity:	37.93%
Best Local Similarity:	21.84%
Query Match:	8.97%
DB:	17
US-10-008..355-2..COPY_522_712 (1-191) x AAH11244 (1-2136)	Length: 2136
QY 2 LysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 21	Matches: 38
DB 1552 AAAAAGCAGACAGTCCGGATGACGAGAGAAAGACAGTAGTCAACGTCGATTAACG 1611	Conservative: 28
QY 22 IleGluIysGlyLysAsnArgLeuPhePheAlaGlyLeuArgGlnMetTyrProGlyArgAla 41	Mismatches: 66
DB 1612 GAAACAATTGGACAAAGATATGTTCTCCCAAGC-----GACGCAACCGATGAAAAAG 1665	Indels: 42
QY 42 LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu 61	Gaps: 7
DB 1666 ATTCACAGCAGCAAAAACATCGTTATCG----- 1695	
QY 62 ProGlnAspGlyAlaTrrpTyrAsnTyr-----HisThrThrGlyLysGly--Val 77	
DB 1696 -----GGGCTTGGTACGGGCGCATATTTGCCAGCAGCAAGCTGACGCGCATAGCT 1746	
QY 78 LeuGluIysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 97	
DB 1747 TCTGATTAAGAGGGCGGCAACAGGCGGGAATTTACTGTG----- 1785	
QY 98 PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeu 117	
DB 1786 -----AATTTGGCGAGAAAAAATAATTACCGGCAAGTTA----- 1818	
QY 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArg 137	
DB 1819 -----ACCCCTGAAAAACAGCAGCAGGCAACCTTACCATTCATGATGTAAG 1863	
QY 138 LeuIleGlyLeuAlaPheAspGlyAsnTrrpGluAlaMetSerGlyAspIleGluPheGlu 157	
DB 1864 ATTGAGGCAACGGTTTTCGCTAGC-----GCAAAAACGTGTAATTAGCTTTTGAT 1917	
QY 158 ProAspLeuGlnArgThrIleSerValAspIleArgTyrVal 171	
DB 1918 CTCGATCAAAAAAATACCACCGCCAGCGCATACAGCATATATC 1959	
RESULT 3	
AAH17774	
ID AAH17774 standard; cDNA; 2794 BP.	
XX AAH17774:	
XX AC	
XX DT 26-JUN-2001 (first entry)	
XX XX	
XX DE Human cDNA sequence SEQ ID NO:17412.	
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss	
XX OS Homo sapiens.	
XX PN EPI074617-A2.	
XX PD 07-FEB-2001.	
XX PF 28-JUL-2000; 2000EP-0116126.	
XX PR 29-JUL-1999; 99JP-0248036.	
XX PR 27-AUG-1999; 99JP-0300253.	
XX PR 11-JAN-2000; 2000JP-0118776.	
XX PR 02-MAY-2000; 2000JP-0183767.	
XX PR 03-JUN-2000; 2000JP-0241899.	

PX	(HELI-) HELIX RES INST.
XX	Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	WPI: 2001-318749/34,
DR	
XX	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
PS	Claim 8; SEQ ID 17A12: 2537pp + CD ROW; English.
XX	
CC	The present invention describes primer sets for synthesising 5602
CC	full-length cDNAs defined in the specification. Where a primer set
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	to the complementary strand of a polynucleotide which comprises one of
CC	the 5602 nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	the specification. The primer sets can be used in antisense therapy and
CC	in gene therapy. The primers are useful for synthesising polynucleotides
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
XX	of the present invention.
XQ	Sequence 2794 BP; 687 A; 851 C; 735 G; 521 T; 0 other:
Alignment Scores:	
Pred. No.:           1 22	Length:         2794
Score:              89.50	Matches:        42
Percent Similarity: 38.32%	Conservative:   22
Best Local Similarity: 25.15%	Mismatches:    56
Query Match:          8.92%	Indels:          47
Bk:                   22	Gaps:            8
US-10-008-355-2_COPY_522_712 (1-191) x AAH17774 (1-2794)	
OY    19 AlAtGyAlalIeGlUlySgLyLysArGLeuPheNealaGlyLeuaRvgLuMeTyrrPro 38 :::    :: :   ::  ::     Db   514 AGCTATGGCGTCACAGAGGGGCCGTGTGCATCAGCATGAATCAATGAGAATACTCC 573 :::    :: :   ::  ::     OY    39 GLyrghlaLeuProSeerAsPaLa-----AsnPhetHeIrmet----- 52 :::    :: :   ::  ::     Db   574 GTGAAGCACCTTGCTACACAGACCCTGACCCCCACAGTTGGCTATCGGCTGGTCCC 633 :::    :: :   ::  ::     OY    53 ---SerTYrGIYSerieILeLYsgLYtTrcLUprOGlnASpslYAIArtPYrAsnTYrhIs 71 :::    :: :   ::  ::     Db   634 GACTTCCTMGACACCCAGAGCTAGGAGGAGACGTTTTCTCC-----TAGGCTATTGGA 684 :::    :: :   ::  ::     OY    72 ThlrTRglYLvsGlgValleUGluSLInASPProlysSerlsrspGUlpheaIVaIGln 91 :::    :: :   ::  ::     Db   685 GGCACTGGGAG-- --AACGCCACAAT----- 708 :::    :: :   ::  ::     OY    92 GlusAnIIeleuNspLeuPheARGrThrlYsaSnTYrgLy---ArgTYrAlagIUasngLY 110 :::    :: :   ::  ::     Db   709 -----AGCCGTTTGAAAACACGAGAACACAAAGTTTGCAAG----- 744 :::    :: :   ::  ::     OY    111 GlnLeuhISIlleaIPhelLeuSerAsnaNsaspIIIEthrGLY----- 124 :::    :: :   ::  ::     Db   745 -----AAGAGTGTGATGGCGTCGCTTGGCGGATTTTGA 777 :::    :: :   ::  ::	

YY	125	--GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe	143
Dd	778	TGTGGAAATGCACGTGGGAACACTCTTTTACCAAGAATGGAAGATGGCATTCCTTC	837
OY	144	AspGlyAsnTrpGluAlaMetSerCylAspIleGluPheGluProAspLeuGlnArgThr	163
Dd	838	CGAATCCAGAAGAAAGCCCTTGCGGGGTCAGGCCCTCTATTCTCATGTCGTGTGAAGAAT	897
OY	164	IleSerValAspIleArgTyr	170
Dd	898	TGCGCAGTGGAGTCAACCTTC	918
RESULT 4			
AAIS9418			
ID	AAIS9418	standard; cDNA; 3259 BP.	
XX	AAIS9418;		
AC			
XX	22-OCT-2001	(first entry)	
DT			
XX			
DE	Human polynucleotide SEQ ID NO 1621.		
XX			
KW	Human; neurotropic; immunosuppressant; cytotstatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KX	chemokineic; thrombolytic; drug screening; arthritis; inflammation;		
KW	Leukaemia; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
PF			
XX	26-DEC-2000; 2000WO-US34263.		
PR			
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HSE-) HSEQ INC.		
PJ			
PJ	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PJ	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
DR	WPI: 2001-442253/47.		
PS	P-PSDB; AAM40262.		
PT			
XX	Novel nucleic acids and polypeptides, useful for treating disorders		
XX	such as central nervous system injuries -		
XX			
PS	Claim 1: SEQ ID NO 1621; 10078pp; English.		
XX			
CC	The invention relates to human nucleic acids (AAIS7798-AAI61369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with neurotropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: immune system suppression.		
CC	activation/inhibit activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		

[illegible]



CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3899 BP; 798 A; 1052 C; 1108 G; 941 T; 0 other;

# Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.92	3899	42	22	56	8
Percent Similarity:	89.50		22	56	47	
Best Local Similarity:	38.32%		56	47		
Query Match:	25.15%		47			
	8.92%					

US-10-008-355-2\_COPY\_522\_712 (1-191) x AAS6370 (1-3899)

QY 19 AlaTyrAlaIleGluLysGlyValArgLeuPhePheAlaGluValArgGluMetTyrPro 38

Db 2982 AGCTATGGGGTCAGAGAGGGCCGCTGATGCTCGAGATGAAGAAATCTCC 2923

QY 39 GAlYArgAlaLeuProSerAspAla-----AsnPhetHmetArgMet----- 52

Db 2922 GTGAGAGACCTTCCTGCTACAGACCTGACCCCGCTGCTATCGGCTGCTCC 2863

QY 53 ---SertYrgLysSerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHis 71

Db 2862 GACCTCCGACAGACCCAGCTAGAGAGCCCTTTC-----TATGCTATGGA 2812

QY 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91

Db 2811 GGCCTCTGGGAG-----AAGTCCACCAAT----- 2788

QY 92 GlnAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110

Db 2787 -----AGCGGTTTGAACACGAGAGCAAGATTGACAG----- 2752

QY 111 GlnLeuHsIleAlaPheLeuSerAsnAspIleThrGly----- 124

Db 2751 -----AAGCATGTGATTTGGCTGCTTTCGCGATTTTGA 2719

QY 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyValGluIleGlyLeuAlaPhe 143

Db 2718 TGTGGAAATGACGTGGAACCTCTTTTACCAAGANTGAAAGATGGATGCTTTC 2659

QY 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163

Db 2658 CGAATCCGAGAGGAAGCCTTGGGGGCTCAGGCCCTCTATCTCATGCTGTAAGAAAT 2599

QY 164 IleSerValAspIleArgTyr 170

Db 2598 TGCAGTGTGAGATTCACTTC 2578

## RESULT 7

AA161204/C  
 ID AA161204 standard; cDNA: 3915 BP.

XX AA161204;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 5193.

XX Human; noctropic; immunosuppressant; cytoskeletal; gene therapy; cancer;

XX peripheral nervous system; neuropathy; central nervous system; CNS;

XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX Chemokine; chromolytic; drug screening; arthritis; inflammation;

XX leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QH, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM42048.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 5193; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

XX the encoded polypeptides (AAM38642-AAM42213) with noctropic,

XX immunosuppressant and cytoskeletal activity. The polynucleotides are useful

XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX of the invention may be used to treat diseases of the peripheral nervous

XX system, such as peripheral nervous injuries, peripheral neuropathy and

XX localised neuropathies and central nervous system diseases, such as

XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX utilisation of the activities such as: Immune system suppression,

XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX assays for receptor activity, arthritis and inflammation, leukaemias and

XX C.N.S disorders.

XX C.N.S disorders.

XX Note: The sequence data for this patent did not form part of the printed

XX specification.

XX Sequence 3915 BP; 798 A; 1041 C; 1122 G; 954 T; 0 other;

XX Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.93	3915	42	22	56	8
Percent Similarity:	89.50		22	56	47	
Best Local Similarity:	38.32%		56	47		
Query Match:	25.15%		47			
	8.92%					

US-10-008-355-2\_COPY\_522\_712 (1-191) x AA161204 (1-3915)

QY 19 AlaTyrAlaIleGluLysGlyValArgLeuPhePheAlaGluValArgGluMetTyrPro 38

Db 2998 AGCTATGGGGTCAGAGAGGGCCGCTGATGCTCGAGATGAAGATGAAGAAATCTCC 2939

QY 39 GAlYArgAlaLeuProSerAspAla-----AsnPhetHmetArgMet----- 52

Db 2938 GTGAGAGACCTTCCTGCTACAGACCTGACCCCGACGAGGCTCCATGCGGCTGCTCC 2879

QY 53 ---SertYrgLysSerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHis 71

Db 2878 GACCTCCTGACAGACCCAGCTAGCGAAGAGCCTTCTCC-----TATGCTATGGA 2828

QY 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91

Db 2827 GGCCTCTGGGAG-----AAGTCCACCAAT----- 2804







QY 19 AlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPro 38  
 Db 4772 GCTTCTCTGTCAGGAATGCAAGAAGTTG----- 4801  
 QY 39 GlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMet---SerTyrGlySerIle 57  
 Db 4802 -----ATTCCATCTTGGGCTTCAGTAAGGAAGACCTTGTGCCCTATGCT----- 4846  
 QY 58 LysGlyTyrGluProGlnAspGlyAlaTriPtyrAsnTyrHis----- 71  
 Db 4847 ---GGCTCATGGAAGTTGGAAGCAGATGGATGGAGAGAGAGGTCAGTTGATCGCG 4903  
 QY 72 ---ThrThrGlyGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaVal 90  
 Db 4904 GCTGTTCCAGGAAGAACGTTGGTCAACGTCACAGACAAAACCCGAGC----- 4948  
 QY 91 GlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGly 110  
 Db 4949 -----TTGTTCAAAGTCAGGAATGGGGGA-----GAAATCGGG 4981  
 QY 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerPro 130  
 Db 4982 GCTGTCGCTCTT-----GACTATCCGAGTGGCACTTCAGGATCTCCT 5023  
 QY 131 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMet 150  
 Db 5024 ATTGTTAACGAACGAGGAGGTGATTGGCTGTACGGCAATGGC-----ATCCTT 5074  
 QY 151 SerGlyAspIleGluPheGluProAspLeuGlnArgThr 163  
 Db 5075 GTCGGTGCACAACCTCTTCGTGTCGGCCATATCCAGACT 5113  
 RESULT 11  
 ID ABA90521  
 AC ABA90521; standard; DNA; 2365589 BP.  
 DT 16-MAY-2002 (first entry)  
 DE Genomic sequence of *Lactococcus lactis* IL1403.  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.  
 OS *Lactococcus lactis* IL1403.  
 PN FR2807446-A1.  
 PD 12-OCT-2001.  
 XX 11-APR-2000; 2000FR-0004630.  
 XX 11-APR-2000; 2000FR-0004630.  
 PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
 PI WPI; 2002-043418/06.  
 DR  
 XX  
 PT New nucleotide sequence useful in the identification of *Lactococcus*  
 PT lactis and related species -  
 XX  
 PS Claim 1; SEQ ID 1; 2504pp; French.

The present invention is related to a *Lactococcus lactis* nucleotide sequence (ABA90521) and related proteins (AB53300-AB55621). The nucleic acid sequence is useful in the detection and/or amplification of nucleic acid sequence, particularly to identify *Lactococcus lactis* or related species. The proteins of the invention are useful for the biosynthesis or biodegradation of a composition of interest. The invention helps research in lactic bacteria, particularly useful in the

CC production of yogurt and cheese.

CC Note: The sequence data for this patent is based on equivalent patent

CC WO200177334 (published 18-OCT-2001) which is available in electronic

CC format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

XX

XX Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

Alignment Scores:

Pred. No.:	3.18e+04	Length:	2365589
Score:	86.00	Matches:	45
Percent Similarity:	36.71%	Conservative:	31
Best Local Similarity:	21.74%	Mismatches:	73
Query Match:	8.57%	Indels:	58
DB:	24	Gaps:	8

US-10-008-355-2\_COPY\_522\_712 (1-191) x ABA90521 (1-2365589)

Qy 14 AspaAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 33

Db 496828 GATCGGTTAGAAAATGGGTTGAAATAGAAAAA----- 496860

Qy 34 ArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThr----- 49

Db 496861 -----CCAAGCATCAATCGATGAGAAAATGGCGATTTGTAAAAAATAAAAG 496908

Qy 50 MetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsn 69

Db 496909 ATTCGTTTAGATTGGCTCAGCTTCAAGGAATTTCTCGTGATTAGCAAAATGGATTGTT 496968

Qy 70 TyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAla 89

Db 496969 GAAATCAACCCCTATAAGATTTTAGCAGACTTTGTCGAAAAATTACCTAATAATTTTCAT 497028

Qy 90 ValGlnGluAsnIleLeuAspLeu-----PheargThrLysAsn 102

Db 497029 AAAAAAGAAACATCTCCCTTTTGATTACAGTTGGAGCCTTGTGATATGCAGATTCAAT 497088

Qy 103 TyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSer-----AsnAsn 120

Db 497089 CGAGGGAATTAGCTTATATTTGGCTGACCATGCTAATTTATTAATTTATAGTAGT 497148

Qy 121 AspIleThrGlyGlyAsnSerGlySerProVal----- 131

Db 497149 GATATTTTATGCAAGTTCTGCTGGGGCTTTGCCATCATCAGGCGAGAAGATTATTCA 497208

Qy 132 -----PheAspLysAsnGlyArgLeuIleGlyLeuAla----- 142

Db 497209 GAAACTGAAAAATACGATTTTGAAAAAAT-----TTATTAGGATTTGGGGTAACACCA 497262

Qy 143 -----PheAspGlyAsnTrpGluAlaMetSerGlyAsp 153

Db 497263 CATCCTTTGCAAAATCTAGCTCCAGATTTGAAGGTAATTTTACACCTCTTGCT----- 497316

Qy 154 IleGluPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPhe 173

Db 497317 --CAACTTGTGAAAAATAGCGAATGACTATCTCTGTTGAAATCAATTACATTAGAACG 497373

Qy 174 MetIleAspLysTrpGlyGln 180

Db 497374 CATCGGACTAAGACAGGGCAA 497394

RESULT 12

ABQ65948

ID ABQ65948 standard; DNA; 651 BP.

XX

XX AC ABQ65948;

XX

XX DF 21-AUG-2002 (first entry)

XX

XX Arabidopsis thaliana polynucleotide SEQ ID NO 525.

XX

KW Arabidopsis thaliana; thale cress; plant; transgenic; GMO; disease;

KW stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;



PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142803.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149388.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
Pred. No.: 0.899 Length: 972  
Score: 85.50 Matches: 42  
Percent Similarity: 38.16% Conservative: 16  
Best Local Similarity: 27.63% Mismatches: 67  
Query Match: 8.52% Indels: 27  
DB: 21 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x AAC46293 (1-972)

QY 45 AspAlaAsnPhetHrMetArgMetSerTyrGlySerIleLysGlyTyrGluPro---Gln 63  
DB 502 GATGCTAAGGGAACAAGATTAGCAAGGAAGGAAATTCGGGTCTTGATCCAGATAAT 561  
QY 64 AspGlyAlaTrpTyrAsnTyrHisThrGlyLys-----Gly 76  
DB 562 GATCTAGCTGTGTGAGATTGAACATGAGGACGTGAACATAATCCTGTTGTCGGT 621  
QY 77 ValLeuGluLysGlnAspProLysSerAspGluPheAlaVal-----Gln 91  
DB 622 ACCTCCAATGATCTACGCGTAGGTCAAGTTGCTTTCGATTGGGAATCCATATGGATAT 681  
QY 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyr-----AlaGluAsn 109  
DB 682 GAAACACCTTGACATAGGGGTAGTAAAGTGGGTGGGAAGAGATACCTTCACCTAAT 741  
QY 110 GlyGln---LeuHisIleAlaPheLeuSerAsnAspIleThrGlyGlyAsnSerGly 128  
DB 742 GGGAGAGTATTTTCAGAGCTATACAAACCGATCTGATATTAACTCAGGCAATTCGTGA 801  
QY 129 SerProValPheAspLysAsnGlyArgLeuLeuGlyLeu-----AlaPheAspGly 145  
DB 802 GGGCATTGCTGATCTTATGCGCATACCATAGGTGTGAACACTGCCACATTCACCCGA 861  
QY 146 AsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSer 165  
DB 862 AAGGGAGTGTATGCTTCTGTTGTTAACTTT-----GCCATTTCCC 903  
QY 166 ValAsp-----IleArgTyrValLeuPheMetIle 175  
DB 904 ATTGACACAGTTGTCGCAACAGTTCCGTCACCTCATT 939

RESULT 14  
AAH53441  
ID AAH53441 standard; DNA; 606 BP.

AC AAH53441;

XX 03-SEP-2001 (first entry)

XX S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2275.  
XX Staphylococcus epidermidis Srl strain; infection; diagnosis;  
XX vaccination; endocarditis; ds.

XX Staphylococcus epidermidis.  
XX WO200134809-A2.  
XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US30782.  
XX 09-NOV-1999; 99US-0164258.  
XX (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
PI WPI; 2001-316495/33.  
DR P-PSDB; AAG82591.  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX Claim 8; Page 605-606; 2188pp; English.  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53871 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4434 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX Sequence 606 BP; 209 A; 92 C; 129 G; 176 T; 0 other;

Alignment Scores:  
Pred. No.: 0.961 Length: 606  
Score: 83.00 Matches: 34  
Percent Similarity: 35.25% Conservative: 15  
Best Local Similarity: 24.46% Mismatches: 46  
Query Match: 8.28% Indels: 44  
DB: 22 Gaps: 6

US-10-008-355-2\_COPY\_522\_712 (1-191) x AAH53441 (1-606)

QY 57 IleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGly 76  
DB 1 ATGAAGGGGGCGCATCCCATGAAACACATCAAGAACCCCTTGTATTAAAGGTAAAGTCA 60  
QY 77 ValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 96  
DB 61 TTGCTAAAGAGTATGATCTTACAGGTGAAGATTT-----GAAGGTCTAATCGAT 111  
QY 97 LeuPheArgThr---LysAsnTyr-----GlyArgTyrAlaGluAsn 109  
DB 112 TTTGCTATGACATTAAAAAATAATAAACACACAGGCACACCATCGATATTTAGAGGGT 171  
QY 110 GlyGlnLeu----- 112  
DB 172 AAGAAATATTGCTTTACTCTTCGAAAAAGACATCTACTCGGACGCGTGGCGCATTTACAGTT 231  
QY 113 -----HisIleAlaPheLeuSerAsnAsnAspIleThrGlyGly 125  
DB 232 GCATCTATTGATCTAGGTGCACACCCCTGAATTTTGGGAAAAAATGATATTCATAGGA 291  
QY 126 AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly 145  
DB 292 AAAAAAGAACTCT-----GTTGAGGATACTGCTAAAGTTTAGGCAGAAATGTTTGAATGA 345

QY 146 AsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164  
 Db 346 -----ATTGAATTAGAGGTTTTTCCCAAAAACGTTT 378  
 RESULT 15  
 AAF91391  
 ID AAF91391 standard; DNA; 731 BP.  
 XX  
 AC AAF91391;  
 DT  
 DT 04-MAY-2001 (first entry)  
 XX  
 XX N. meningitidis (serogroup B) TbpA gene upstream sequence, SEQ ID:17.  
 XX  
 XX Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;  
 KW genetically modified; protective antigen expression; LPS detoxification;  
 KW LPS; lipid A; homologous recombination vector; immunisation;  
 KW immunoprotective; non-toxic; paediatric; cyclic; circular; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 XX WO200109350-A2.  
 PN  
 PN 08-FEB-2001.  
 PD  
 XX 31-JUL-2000; 2000WO-EP07424.  
 XX  
 XX 03-AUG-1999; 99GB-0018319.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 XX Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;  
 PI Poolman J, Thirry G, Thonnard J, Voet P;  
 XX  
 XX WPI; 2001-138654/14.  
 DR  
 XX  
 PT New isolated polynucleotide useful for outer membrane vesicle  
 PT preparation from Gram-negative bacterial strain for vaccination of  
 PT microbial infections -  
 XX  
 PS Claim 46; Page 81; 128pp; English.  
 XX  
 CC The invention relates to a genetically-engineered outer membrane vesicle  
 CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.  
 CC The blebs of the invention are improved with respect to their  
 CC immunogenicity and toxicity by the introduction of one or more genetic  
 CC changes to the chromosome of the bacterium from which the blebs are  
 CC derived. The changes made include the upregulation of protective antigen  
 CC expression, the downregulation of immunodominant non-protective antigen  
 CC expression, and genetic changes which result in detoxification of the  
 CC Lipid A moiety of lipopolysaccharide (LPS). The invention also  
 CC encompasses modified Gram-negative bacterial strains from which the bleb  
 CC preparations are made, a vector suitable for performing recombination  
 CC events (for the generation of the modified bacterial strains),  
 CC bacterially-derived nucleic acid sequences used in such a vector, and an  
 CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole  
 CC cell vaccine suitable for paediatric use. The bleb preparation is useful  
 CC in the manufacture of a medicament for immunising a human host against a  
 CC disease caused by infection of one or more of the following: Neisseria  
 CC meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella  
 CC catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia  
 CC pneumoniae. The invention may also be used to provide immunisation against  
 CC the influenza virus. Bacterially derived nucleotide sequences of the  
 CC invention are used in the performance of homologous recombination events  
 CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either  
 CC increase or decrease expression of that gene. Immunoprotective and  
 CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines  
 CC are more immunogenic, less toxic and safer, and are particularly useful  
 CC for paediatric use. The present sequence represents a specifically  
 CC claimed Neisseria meningitidis nucleic acid sequence.  
 XX  
 SQ Sequence 731 BP; 226 A; 169 C; 194 G; 142 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.24 Length: 731  
 Score: 83.00 Matches: 37  
 Percent Similarity: 35.06% Conservative: 24  
 Best Local Similarity: 21.26% Mismatches: 71  
 Query Match: 8.28% Indels: 42  
 DB: 22 Gaps: 6  
 US-10-008-355-2\_COPY\_522\_712 (1-191) x AAF91391 (1-731)  
 QY 2 LysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 21  
 Db 61 AAAAACAAGCAAGTCGGGATGCGAGGAGCAAGACAGCTAGCTCAAGCTGCTAAAACG 120  
 QY 22 IleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAla 41  
 Db 121 GAACAGTTGGACAAATATGTTCTCCACAGC-----GAGCGCAGCGATGAAAGAG 174  
 QY 42 LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu 61  
 Db 175 ATCCAAACGACCAAAAGCTGCTTTATCGG----- 204  
 QY 62 ProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLys-----GlyVal 77  
 Db 205 -----GGGTCTTGGTACGGGCATATTGCCAACGCGACCAAGCTGGAGCGCAATGCT 255  
 QY 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 97  
 Db 256 TCCGATAAAGAGGGCGGCAACAGGGCGGACTTTACTGTG----- 294  
 QY 98 PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeu 117  
 Db 295 -----AATTTCCGTACGAAAAAATTAACGCGCAGTTA----- 327  
 QY 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArg 137  
 Db 328 -----ACCGCTGACACACAGCGCGGCGCAACCTTTACCATTGCTGGCGCAT 372  
 QY 138 LeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGlu 157  
 Db 373 ATTGAGGCAACGGTTTTTCCGGTACG-----CGAANAACCTGCTCACTCAGGTTTGTAT 426  
 QY 158 ProAspLeuGlnArgThrIleSerValAspIleArgTyrVal 171  
 Db 427 CTCGATCAAAAGCAATAACACCCGCGCGCTAAGGCATATATC 468

Search completed: May 23, 2003, 10:12:49  
 Job time : 506.818 secs





GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 08:43:01 ; Search time 1054.2 seconds  
(without alignments)  
5272.840 Million cell updates/sec

Title: US-10-008-355-2\_COPY\_522\_712

Perfect score: 1003

Sequence: 1 SKSVIAARAIAQADANAY.....LFMDKWGQCPLRIQLKLI 191

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spool/US1008355/runat\_16052003\_110400\_2535/app\_query.fasta\_1.1230  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US1008355@cgn\_1.1\_2983@runat\_16052003\_110400\_2535 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_em.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_mu.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	334	33.3	10689	1	AE004008	AE004008 Xylella f
c 2	176.5	17.6	253217	2	AC016590	AC016590 Homo sapi
c 3	175	17.4	10811	1	AE011732	AE011732 Xanthomon
c 4	101.5	10.1	204050	1	AL646073	AL646073 Ralstonia
c 5	93	9.3	11581	1	AE007581	AE007581 Clostridi
c 6	91.5	9.1	79718	8	AF012H2	AF012H2 Aspergill
c 7	91.5	9.1	180688	9	CNS01RG9	AL157871 Human chr
c 8	91.5	9.1	187710	9	CNS01DVI	AL135838 Human chr
c 9	90.5	9.0	1554	14	AIU52600	U52600 Avian infec
c 10	90	9.0	2143	6	A47465	A47465 Sequence 38
c 11	90	9.0	2163	1	NMTBP2163	250731 N.meningiti
c 12	89.5	8.9	2794	9	AK023006	AK023006 Homo sapi
c 13	89.5	8.9	2908	10	BC021506	BC021506 Mus muscu
c 14	89.5	8.9	2956	9	BC027713	BC027713 Homo sapi
c 15	89.5	8.9	3048	9	BC009988	BC009988 Homo sapi
c 16	89.5	8.9	3058	10	BC027844	BC027844 Mus muscu
c 17	89.5	8.9	3158	9	BC002564	BC002564 Homo sapi
c 18	89.5	8.9	3513	9	HS07509	AK007509 Homo sapi
c 19	89.5	8.9	3514	9	AK022863	AK022863 Homo sapi
c 20	89.5	8.9	10862	6	AX124177	AX124177 Sequence
c 21	89.5	8.9	10862	6	AX002387	AX002387 Sequence
c 22	89.5	8.9	10862	14	AF052437	AF052437 Yellow fe
c 23	89.5	8.9	10862	14	AF052438	AF052438 Yellow fe
c 24	89.5	8.9	10862	14	AF052439	AF052439 Yellow fe
c 25	89.5	8.9	10862	14	AF052444	AF052444 Yellow fe
c 26	89.5	8.9	10862	14	AF052445	AF052445 Yellow fe
c 27	89.5	8.9	10862	14	AF052446	AF052446 Yellow fe
c 28	89.5	8.9	10862	14	FLYF17DG	X03700 Yellow feve
c 29	89.5	8.9	10862	14	YFUI17066	U17066 Yellow feve
c 30	89.5	8.9	10862	14	YFUI17067	U17067 Yellow feve
c 31	89.5	8.9	10862	14	YFUI17068	U17068 Yellow feve
c 32	89.5	8.9	10862	14	YFUI17069	U17069 Yellow feve
c 33	89.5	8.9	10862	14	YFUI17070	U17070 Yellow feve
c 34	89.5	8.9	10862	14	YFUI17071	U17071 Yellow feve
c 35	88	8.8	12602	3	DEU40653	U40653 Drosophila
c 36	87	8.7	315079	1	MPULM03	AL445565 Mycoplasma
c 37	86.5	8.6	3832	1	CAJFLAAB	J05635 Campylobact
c 38	86.5	8.6	7756	1	AF202168	AF202168 Campyloba
c 39	86.5	8.6	10760	14	AF094612	AF094612 Yellow fe
c 40	86	8.6	9293	1	AE011022	AE011022 Methanosa
c 41	86	8.6	10590	1	AE006285	AE006285 Lactococc
c 42	85.5	8.5	1003	8	AY091427	AY091427 Arabidops
c 43	85.5	8.5	1156	8	AF114386	AF114386 Arabidops
c 44	85.5	8.5	1177	8	AY056227	AY056227 Arabidops
c 45	84.5	8.4	1731	1	AF050190	AF050190 Campyloba

ALIGNMENTS

RESULT 1

AE004008/c	AE004008	10689 bp	DNA	linear	BCT 15-JUN-2001
LOCUS	Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.				
DEFINITION	Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.				
ACCESSION	AE004008 AE003849				
VERSION	AE004008.1 GI:9106961				
KEYWORDS					
SOURCE	Xylella fastidiosa 9a5c.				
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.				
REFERENCE	1 (bases 1 to 10689)				
AUTHORS	Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvaranga,R., Alves,L.M., Araya,J.E., Baig,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorfy,H., Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P. and Marino,C.L.				
TITLE	The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis				
JOURNAL	Nature 406 (6792), 151-157 (2000)				
MEDLINE	20365717				
PUBMED	10910347				
REFERENCE	2 (bases 1 to 10689)				
AUTHORS	Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvaranga,R., Alves,L.M.C., Araya,J.E., Baig,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorfy,H., Facincani,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.				
Direct Submission					
Submitted (02-JUN-2000)	Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil				
FEATURES	Location/Qualifiers				
source	1..10689				
	/organism="Xylella fastidiosa 9a5c"				
	/db_xref="taxon:160492"				
	/clone="9a5c"				
gene	complement(119..367)				
	/gene="XF1871"				
CDS	complement(119..367)				
	/gene="XF1871"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84677.1"				
	/db_xref="GI:9106962"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVNAHTOKL MKETLKTTSKWYPVVGSSLLIAGATAATLFIKLFH"				
	complement(430..618)				
	/gene="XF1872"				
	complement(430..618)				
	/gene="XF1872"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84678.1"				
	/db_xref="GI:9106963"				
	/translation="MMEEARATAPQGGLAASDPQRTIDAPKKTGLRKRKYPTTGSASI IVLRHVLTORQQRKASP"				
	941..1288				
	/gene="XF1873"				
	941..1288				
	/gene="XF1873"				
	/note="similar to GI:479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAF84679.1"				
	/db_xref="GI:9106964"				
	/translation="MSKNKIPNOILTNWAGFOFINGKLVTPERRTIEEWLRWLSLT CTLARWQKNMEARATAPQGGLAASDPQRTIDALQKTVSHKRHAANETASIIQLRQ VLIQROQKRKASP"				
	1351..1662				
	/gene="XF1874"				
	1351..1662				
	/gene="XF1874"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84680.1"				
	/db_xref="GI:9106965"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVTHQTEM NAHINEMNARTQEMNARTQKLIETQKVSQWYPPVVASGLLAGATAATLFIKLF H"				
	complement(1839..2918)				
	/gene="XF1875"				
	complement(1839..2918)				
	/gene="XF1875"				
	/note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="phage-related protein"				
	/protein_id="AAF84681.1"				
	/db_xref="GI:9106966"				
	/translation="MLNLVTGAPNGKTLVAVDWLIRQIEIDKSLVKSGAVPRSFYTD IEGDFEAVRLTGYVVSAPEDWRTTPQGSVIVYDEAHMFAGRGRSDDPRVCDL DTHRGYDLMFVYQWPTKTHHELRLVGEHILNRAMGLQTAGLYRWSRAQDDPYDI HQREKDEEYWKFPKRYALYASLTHVTSKHKFRPKKYNALSVCYTCVIGLVFMI YYSPAHLSAASSVAGASGOASLAPASLSSSLVSGMRVTAVLETSAPTLSGCV SSSRSCRCFNTDGYQIDMSVVECPRLASLPFNVYHAYVTSSSSAAAVSSSPSSPS SSVSLSSSVSSAPILGSSSVPSH"				
	complement(2918..3202)				
	/gene="XF1876"				
	complement(2918..3202)				
	/gene="XF1876"				
	/note="similar to SPIP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84677.1"				
	/db_xref="GI:9106962"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVNAHTOKL MKETLKTTSKWYPVVGSSLLIAGATAATLFIKLFH"				
	complement(430..618)				
	/gene="XF1872"				
	complement(430..618)				
	/gene="XF1872"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84678.1"				
	/db_xref="GI:9106963"				
	/translation="MMEEARATAPQGGLAASDPQRTIDAPKKTGLRKRKYPTTGSASI IVLRHVLTORQQRKASP"				
	941..1288				
	/gene="XF1873"				
	941..1288				
	/gene="XF1873"				
	/note="similar to GI:479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAF84679.1"				
	/db_xref="GI:9106964"				
	/translation="MSKNKIPNOILTNWAGFOFINGKLVTPERRTIEEWLRWLSLT CTLARWQKNMEARATAPQGGLAASDPQRTIDALQKTVSHKRHAANETASIIQLRQ VLIQROQKRKASP"				
	1351..1662				
	/gene="XF1874"				
	1351..1662				
	/gene="XF1874"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84680.1"				
	/db_xref="GI:9106965"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVTHQTEM NAHINEMNARTQEMNARTQKLIETQKVSQWYPPVVASGLLAGATAATLFIKLF H"				
	complement(1839..2918)				
	/gene="XF1875"				
	complement(1839..2918)				
	/gene="XF1875"				
	/note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="phage-related protein"				
	/protein_id="AAF84681.1"				
	/db_xref="GI:9106966"				
	/translation="MLNLVTGAPNGKTLVAVDWLIRQIEIDKSLVKSGAVPRSFYTD IEGDFEAVRLTGYVVSAPEDWRTTPQGSVIVYDEAHMFAGRGRSDDPRVCDL DTHRGYDLMFVYQWPTKTHHELRLVGEHILNRAMGLQTAGLYRWSRAQDDPYDI HQREKDEEYWKFPKRYALYASLTHVTSKHKFRPKKYNALSVCYTCVIGLVFMI YYSPAHLSAASSVAGASGOASLAPASLSSSLVSGMRVTAVLETSAPTLSGCV SSSRSCRCFNTDGYQIDMSVVECPRLASLPFNVYHAYVTSSSSAAAVSSSPSSPS SSVSLSSSVSSAPILGSSSVPSH"				
	complement(2918..3202)				
	/gene="XF1876"				
	complement(2918..3202)				
	/gene="XF1876"				
	/note="similar to SPIP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84677.1"				
	/db_xref="GI:9106962"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVNAHTOKL MKETLKTTSKWYPVVGSSLLIAGATAATLFIKLFH"				
	complement(430..618)				
	/gene="XF1872"				
	complement(430..618)				
	/gene="XF1872"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84678.1"				
	/db_xref="GI:9106963"				
	/translation="MMEEARATAPQGGLAASDPQRTIDAPKKTGLRKRKYPTTGSASI IVLRHVLTORQQRKASP"				
	941..1288				
	/gene="XF1873"				
	941..1288				
	/gene="XF1873"				
	/note="similar to GI:479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAF84679.1"				
	/db_xref="GI:9106964"				
	/translation="MSKNKIPNOILTNWAGFOFINGKLVTPERRTIEEWLRWLSLT CTLARWQKNMEARATAPQGGLAASDPQRTIDALQKTVSHKRHAANETASIIQLRQ VLIQROQKRKASP"				
	1351..1662				
	/gene="XF1874"				
	1351..1662				
	/gene="XF1874"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84680.1"				
	/db_xref="GI:9106965"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVTHQTEM NAHINEMNARTQEMNARTQKLIETQKVSQWYPPVVASGLLAGATAATLFIKLF H"				
	complement(1839..2918)				
	/gene="XF1875"				
	complement(1839..2918)				
	/gene="XF1875"				
	/note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="phage-related protein"				
	/protein_id="AAF84681.1"				
	/db_xref="GI:9106966"				
	/translation="MLNLVTGAPNGKTLVAVDWLIRQIEIDKSLVKSGAVPRSFYTD IEGDFEAVRLTGYVVSAPEDWRTTPQGSVIVYDEAHMFAGRGRSDDPRVCDL DTHRGYDLMFVYQWPTKTHHELRLVGEHILNRAMGLQTAGLYRWSRAQDDPYDI HQREKDEEYWKFPKRYALYASLTHVTSKHKFRPKKYNALSVCYTCVIGLVFMI YYSPAHLSAASSVAGASGOASLAPASLSSSLVSGMRVTAVLETSAPTLSGCV SSSRSCRCFNTDGYQIDMSVVECPRLASLPFNVYHAYVTSSSSAAAVSSSPSSPS SSVSLSSSVSSAPILGSSSVPSH"				
	complement(2918..3202)				
	/gene="XF1876"				
	complement(2918..3202)				
	/gene="XF1876"				
	/note="similar to SPIP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84677.1"				
	/db_xref="GI:9106962"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVNAHTOKL MKETLKTTSKWYPVVGSSLLIAGATAATLFIKLFH"				
	complement(430..618)				
	/gene="XF1872"				
	complement(430..618)				
	/gene="XF1872"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84678.1"				
	/db_xref="GI:9106963"				
	/translation="MMEEARATAPQGGLAASDPQRTIDAPKKTGLRKRKYPTTGSASI IVLRHVLTORQQRKASP"				
	941..1288				
	/gene="XF1873"				
	941..1288				
	/gene="XF1873"				
	/note="similar to GI:479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAF84679.1"				
	/db_xref="GI:9106964"				
	/translation="MSKNKIPNOILTNWAGFOFINGKLVTPERRTIEEWLRWLSLT CTLARWQKNMEARATAPQGGLAASDPQRTIDALQKTVSHKRHAANETASIIQLRQ VLIQROQKRKASP"				
	1351..1662				
	/gene="XF1874"				
	1351..1662				
	/gene="XF1874"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84680.1"				
	/db_xref="GI:9106965"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVTHQTEM NAHINEMNARTQEMNARTQKLIETQKVSQWYPPVVASGLLAGATAATLFIKLF H"				
	complement(1839..2918)				
	/gene="XF1875"				
	complement(1839..2918)				
	/gene="XF1875"				
	/note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="phage-related protein"				
	/protein_id="AAF84681.1"				
	/db_xref="GI:9106966"				
	/translation="MLNLVTGAPNGKTLVAVDWLIRQIEIDKSLVKSGAVPRSFYTD IEGDFEAVRLTGYVVSAPEDWRTTPQGSVIVYDEAHMFAGRGRSDDPRVCDL DTHRGYDLMFVYQWPTKTHHELRLVGEHILNRAMGLQTAGLYRWSRAQDDPYDI HQREKDEEYWKFPKRYALYASLTHVTSKHKFRPKKYNALSVCYTCVIGLVFMI YYSPAHLSAASSVAGASGOASLAPASLSSSLVSGMRVTAVLETSAPTLSGCV SSSRSCRCFNTDGYQIDMSVVECPRLASLPFNVYHAYVTSSSSAAAVSSSPSSPS SSVSLSSSVSSAPILGSSSVPSH"				
	complement(2918..3202)				
	/gene="XF1876"				
	complement(2918..3202)				
	/gene="XF1876"				
	/note="similar to SPIP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84677.1"				
	/db_xref="GI:9106962"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVNAHTOKL MKETLKTTSKWYPVVGSSLLIAGATAATLFIKLFH"				
	complement(430..618)				
	/gene="XF1872"				
	complement(430..618)				
	/gene="XF1872"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84678.1"				
	/db_xref="GI:9106963"				
	/translation="MMEEARATAPQGGLAASDPQRTIDAPKKTGLRKRKYPTTGSASI IVLRHVLTORQQRKASP"				
	941..1288				
	/gene="XF1873"				
	941..1288				
	/gene="XF1873"				
	/note="similar to GI:479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAF84679.1"				
	/db_xref="GI:9106964"				
	/translation="MSKNKIPNOILTNWAGFOFINGKLVTPERRTIEEWLRWLSLT CTLARWQKNMEARATAPQGGLAASDPQRTIDALQKTVSHKRHAANETASIIQLRQ VLIQROQKRKASP"				
	1351..1662				
	/gene="XF1874"				
	1351..1662				
	/gene="XF1874"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84680.1"				
	/db_xref="GI:9106965"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVTHQTEM NAHINEMNARTQEMNARTQKLIETQKVSQWYPPVVASGLLAGATAATLFIKLF H"				
	complement(1839..2918)				
	/gene="XF1875"				
	complement(1839..2918)				
	/gene="XF1875"				
	/note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="phage-related protein"				
	/protein_id="AAF84681.1"				
	/db_xref="GI:9106966"				
	/translation="MLNLVTGAPNGKTLVAVDWLIRQIEIDKSLVKSGAVPRSFYTD IEGDFEAVRLTGYVVSAPEDWRTTPQGSVIVYDEAHMFAGRGRSDDPRVCDL DTHRGYDLMFVYQWPTKTHHELRLVGEHILNRAMGLQTAGLYRWSRAQDDPYDI HQREKDEEYWKFPKRYALYASLTHVTSKHKFRPKKYNALSVCYTCVIGLVFMI YYSPAHLSAASSVAGASGOASLAPASLSSSLVSGMRVTAVLETSAPTLSGCV SSSRSCRCFNTDGYQIDMSVVECPRLASLPFNVYHAYVTSSSSAAAVSSSPSSPS SSVSLSSSVSSAPILGSSSVPSH"				
	complement(2918..3202)				
	/gene="XF1876"				
	complement(2918..3202)				
	/gene="XF1876"				
	/note="similar to SPIP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84677.1"				
	/db_xref="GI:9106962"				
	/translation="MTMTLKEDKTEIIDFQRIQIAKLLSETAKIQAEIHVEVNAHTOKL MKETLKTTSKWYPVVGSSLLIAGATAATLFIKLFH"				
	complement(430..618)				
	/gene="XF1872"				
	complement(430..618)				
	/gene="XF1872"				
	/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAF84678.1"				
	/db_xref="GI:9106963"				
	/translation="MMEEARATAPQGGLAASDPQRTIDAPKKTGLRKRKYPTTGSASI IVLRHVLTORQQRKASP"				
	941..1288				
	/gene="XF1873"				
	941..1288				
	/gene="XF1873"				
	/note="similar to GI:479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAF84679.1"				
	/db_xref="GI:9106964"				
	/translation="MSKNKIPNOILTNWAGFOFINGKLVTPERRTIEEWLRWLSLT CTLARWQKNMEARATAPQGGLAASDPQRTIDALQKTVSHKRHAANETASIIQLRQ VLIQROQKRKASP"				
	1351..1662				
	/gene="XF1874"				
	1351..1662</				

QY	48	pheThrMetArgMetSerTyrGlySerIleuysGlyTyrGluProGlnAspGlyAlaTrp	67
Db	8789	TCATCATACGTATTACCTTCGCCACGCTGAAGAGCTATTACACCCAAAGATGGCGTCGAG	8730
QY	68	TyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGlu	87
Db	8729	TACACACCGCTTTACGACACTGCAGGTGCTATGGCGAAG--AATACAGGCGTCGAGCCA	8673
QY	88	PheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAla	107
Db	8672	TTGCATTACCGAAGTCACTGATTAATCGGATAAAGGCAAGAGCTACGCCAATTAGCC	8613
QY	108	GluAsn-----GlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGly	125
Db	8612	GATCAGCGATTGGCAGCTGCTCCCTCTGAATTTCTATCTGACTTGGACATTACTGCGCGC	8553
QY	126	AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly	145
Db	8552	AACCTCGGATGCCAGCTCTAGATCCCAAGAAATTTGGTGGTTTAGCATTTTTGACGGT	8493
QY	146	AsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSer	165
Db	8492	AATTGGGAGTGTCTAGTAGTAATTTGGGTGTTTGATCTGTGATGACTCGAACAATTGCA	8433
QY	166	ValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeuIle	185
Db	8432	GTTGATAGCTGTATGTGTCAGTGGATCATGACTGAAGTTCACAGCACCATCTACTACTA	8373
QY	186	GlnGluLeuLysLeu	190
Db	8372	AAGGAATTGAATTTG	8358

RESULT 2

AC016590/c

LOCUS

DEFINITION

SEQUENCE, 39 unordered pieces.

ACCESSION

AC016590

VERSION

AC016590.6

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 253217)

DOE Joint Genome Institute.

AUTHORS

Sequencing of Human Chromosome 19

TITLE

Unpublished

REFERENCE

2 (bases 1 to 253217)

DOE Joint Genome Institute.

AUTHORS

Direct Submission

TITLE

Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

JOURNAL

On Apr 20, 2001 this sequence version replaced gi:7711568.

COMMENT

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

-----

Project Information

Center Project Name: 940643, BC905667

Center clone name: CITB-EL\_3220F14

-----

Summary Statistics

Consensus quality: 207648 bases at least Q40

Consensus quality: 225569 bases at least Q30

Consensus quality: 232428 bases at least Q20

Estimated insert size: 250510; agarose-fp estimation

Estimated insert size: 249417; sum-of-contigs estimation

Quality coverage: 9.03 in Q20 bases; agarose-fp estimation

Quality coverage: 9.07 in Q20 bases; sum-of-contigs estimation.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 39 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1019: contig of 1019 bp in length  
\* 1020 1119: gap of unknown length  
\* 1120 2152: contig of 1033 bp in length  
\* 2153 2252: gap of unknown length  
\* 2253 3701: contig of 1449 bp in length  
\* 3702 3801: gap of unknown length  
\* 3802 5153: contig of 1252 bp in length  
\* 5154 5154: gap of unknown length  
\* 5154 6184: contig of 1031 bp in length  
\* 6185 6284: gap of unknown length  
\* 6285 7809: contig of 1525 bp in length  
\* 7810 7909: gap of unknown length  
\* 7910 8986: contig of 1077 bp in length  
\* 8987 9086: gap of unknown length  
\* 9087 10090: contig of 1004 bp in length  
\* 10091 10190: gap of unknown length  
\* 10191 11225: contig of 1035 bp in length  
\* 11226 11325: gap of unknown length  
\* 11326 12365: contig of 1040 bp in length  
\* 12366 12465: gap of unknown length  
\* 12466 14535: contig of 2070 bp in length  
\* 14536 14635: gap of unknown length  
\* 14636 16489: contig of 1854 bp in length  
\* 16490 16589: gap of unknown length  
\* 16590 19109: contig of 2520 bp in length  
\* 19110 19209: gap of unknown length  
\* 19111 21841: contig of 2632 bp in length  
\* 21842 21941: gap of unknown length  
\* 21942 23735: contig of 1794 bp in length  
\* 23736 23835: gap of unknown length  
\* 23836 25861: contig of 2026 bp in length  
\* 25862 25961: gap of unknown length  
\* 25962 29117: contig of 3156 bp in length  
\* 29118 29217: gap of unknown length  
\* 29219 31907: contig of 2690 bp in length  
\* 31908 32007: gap of unknown length  
\* 32008 36364: contig of 4357 bp in length  
\* 36365 36464: gap of unknown length  
\* 36465 43309: contig of 6845 bp in length  
\* 43310 43409: gap of unknown length  
\* 43410 48185: contig of 4776 bp in length  
\* 48186 48285: gap of unknown length  
\* 48286 55626: contig of 7341 bp in length  
\* 55627 55726: gap of unknown length  
\* 55727 65781: contig of 10055 bp in length  
\* 65782 65881: gap of unknown length  
\* 65882 69673: contig of 3792 bp in length  
\* 69674 69773: gap of unknown length  
\* 69774 73389: contig of 9616 bp in length  
\* 73390 79489: gap of unknown length  
\* 79490 91453: contig of 11964 bp in length  
\* 91454 91553: gap of unknown length  
\* 91554 96671: contig of 5118 bp in length  
\* 96672 96771: gap of unknown length  
\* 96772 100544: contig of 3773 bp in length  
\* 100545 100644: gap of unknown length  
\* 100645 111423: contig of 10779 bp in length  
\* 111424 111523: gap of unknown length  
\* 111524 115278: contig of 3755 bp in length  
\* 115279 124011: gap of unknown length  
\* 124012 124111: contig of 8633 bp in length  
\* 124112 138475: contig of 14364 bp in length  
\* 138476 138575: gap of unknown length  
\* 138576 140613: contig of 2038 bp in length  
\* 140614 140713: gap of unknown length

140714 144949: contig of 4236 bp in length  
\* 144950 145049: gap of unknown length  
\* 145050 164352: contig of 19303 bp in length  
\* 164353 164452: gap of unknown length  
\* 164453 170540: contig of 6088 bp in length  
\* 170541 170541: gap of unknown length  
\* 170541 185046: contig of 14406 bp in length  
\* 185047 185047: gap of unknown length  
\* 185047 222209: contig of 37063 bp in length  
\* 222210 222309: gap of unknown length  
\* 222310 253217: contig of 30908 bp in length.  
\* 253217 Location/Qualifiers  
1. 253217  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="CTD-3220F14"  
/clone\_lib="Caltech human BAC library D"  
BASE COUNT 62246 a 62867 c 62532 g 61752 t 3820 others  
ORIGIN

Alignment Scores:  
Pred. No.: 3.94e-07 Length: 253217  
Score: 176.50 Matches: 47  
Percent Similarity: 54.19% Conservative: 37  
Best Local Similarity: 30.32% Mismatches: 66  
Query Match: 17.60% Indels: 5  
DB: 2 Gaps: 2

US-10-008-355-2\_COPY\_522\_712 (1-191) x AC016590 (1-253217)  
Qy 39 GlyArgAlaLeuProSerAspAlaAsnPhetThrMetArgMetSerTyrGlySerIleLys 58  
Db 4723 GGCACAGCCGCTATCTCTGACGCCCAACAGCACCTCGCGGTGACCTCGCGCGGTGCC 4664  
Qy 59 GlyTyrGluPro---GlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyVal 77  
Db 4663 GGCGGTGACCATTCGCGCGGACGCCACCTGGACCGCGGTTCACACCGCTCAATGGCGTG 4604  
Qy 78 LeuGluLysGlnAspProLys-SerAspGluPheAlaValGlnGluAsnIleLeuAspLe 97  
Db 4603 GTAGCCAAAGCCAGCGCGGAGGAGTTCATGACCCAGCAACACTGGCGGTGGA 4544  
Qy 97 uPheArgThrLysAsnTyrGlyArgTyrAlaGluAsn-----GlyGlnLeuHisIleAl 115  
Db 4543 TCATGGCCTAAGGACTTCGCGCAATTTGTGATCCGAAGCTGAAACCGTGGCGGTGGA 4484  
Qy 115 aPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAs 135  
Db 4483 CTACTGGCCACCTGGACATCAGCGCGGCACTCCCGGTTCCGCTGCGCTGAACTCGAA 4424  
Qy 135 nGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleG1 155  
Db 4423 AGGCAATTCATCGCGCTGGCGTTCGACGCGCACCTCGATCGATCATTTCCGACTGGGA 4364  
Qy 155 uPheGluPro-AspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetI 175  
Db 4363 CTACAAACAGCCGCAACACCGCTTCGATCCAGTCCAGTCCGCTGATCATTTCCGACTGGGA 4304  
Qy 175 leAspLysTrpGlyGlnCysProArgLeuIleGlnGluLeu 188  
Db 4303 TGAACACGTCGACCCAGCAGCAACCTGCTGAAAGAAATG 4263

RESULT 3  
AE011732 10811 bp DNA linear BCT 29-MAY-2002  
LOCUS Xanthomonas axonopodis pv. citri str. 306, section 110 of 469 of  
DEFINITION the complete genome.  
ACCESSION AE011732 AE008923  
VERSION AE011732.1 GI:21107161  
KEYWORDS Xanthomonas axonopodis pv. citri str. 306.  
SOURCE Xanthomonas axonopodis pv. citri str. 306  
ORGANISM



```

gene      DGFNAVRPADELSAGPSARIYVLDATDVTAPAKAAMAPVRR"
7051..7236
/gene="XAC1027"
CDS       7051..7236
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA035910.1"
/db_xref="GI:21107167"
MTEVAPAPQLLKELGVR"
7329..7973
/gene="pgmA"
/note="XAC1028"
7329..7973
/gene="pgmA"
/note="XAC1029"
8242..9540
/gene="folC"
/note="XAC1029"
8242..9540
/gene="folC"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="phosphoglycerate mutase"
/protein_id="AA035911.1"
/db_xref="GI:21107168"
LOIDRAVASPLRAQATAKAALCASQALLODADLQEIAGHEWGLLASEINDKQPA
RLRMAREEDPTVLPMGESLROVLDRSGLMRAADGLGAHDITLLVVAHDVNRVILC
KIIGLPLSRMSFRQAPTLLNLEGDDVEHLEVVRLNDCAHHTFFFGKAKHRL"
8242..9540
/gene="folC"
/note="XAC1029"
8242..9540
/gene="folC"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark; dihydrofolate
synthase"
/codon_start=1
/transl_table=11
/product="folylpolyglutamate synthase"
/protein_id="AA035912.1"
/db_xref="GI:21107169"
/transl_table="MNATDSLAWLAYIEQQHPSAIAIANGLERVREVAVRVLRQIAAPAKH
VIVVGTNGRSTVAFTEAIGRAAGWKVGATSPHLIRYNERVIDQEASDQQLVDA
FAAVERAGDTALTYPFGTLLAALWLFQSSALELAVLEIGLGRGLDAVNIVDSVAVI

Alignment Scores:
Pred. No.:      1.27e-08      Length:      10811
Score:          175.00      Matches:      32
Percent Similarity: 75.00%      Conservative: 16
Best Local Similarity: 50.00%      Mismatches: 16
Query Match:     17.45%      Indels:      0
DB:              1          Gaps:          0

US-10-008-355-2_COPY_522_712 (1-191) x AE011732 (1-10811)
Oy 125 GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuLeuGlyLeuAlaPheAsp 144
Db 7033 GCGAACTCCGGTTCGCCGTGATGGACCGCAGCGAGCTGGTCCGGCTCGCATC 7092
Oy 145 GlyAsnTrpGluAlaMetSerGlyAspTleGluPheGluProAspLeuGlnArgThrIle 164
Db 7093 GGAATTTGGAGTCCGGTACGACGACCACTGGATCTTCGACCGCAATGACCCGCGATGTC 7152
Oy 165 SerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeu 184
Db 7153 GCCGTGGAGCGCCGCTATCTCGTGGATCATGACCGAGAGTGGCTCCGCGCCACAGCTG 7212
Oy 185 IleGlnGluLeu 188
Db 7213 TTGAAGGAACATC 7224

```

```

RESULT 4
AL646073/c
LOCUS
DEFINITION Ralstonia solanacearum GM11000 chromosome, complete sequence;
segment 17/19.
ACCESSION AL646073 AL646052
VERSION AL646073.1 GI:17429991
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
REFERENCE 1 (bases 1 to 204050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
Siquier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
REFERENCE 2 (bases 1 to 204050)
AUTHORS Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMG CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genétique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
1..204050
/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
complement(115..1278)
/gene="dgt"
/note="RSC2968; RS01329"
complement(115..1278)
/gene="dgt"
/EC_number="3.1.5.1"
/function="small molecule metabolism; central intermediary
metabolism; nucleotide hydrolysis"
/note="product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE DEOXYGUANOSINE TRIPHOSPHATE
TRIPHOSPHOHYDROLASE. PROTEIN"
/protein_id="CAD16677.1"
/db_xref="GI:17429992"
/transl_table="MNDRLDLLPAPFLEGLHAPYAAHAQSGRGRVHAIEPPSTSRTEF
QRDRRIIHCFTAFRRLEYKTVFVFNHEGDLFRTLRLTHSLVLAQIARSTARSRLRNEDL
VEATSLADHLGHTPFHAGQDALNDCKMPYGGFHNQSLVVDRLERYGGFQGLNL
TFETREGILKHCSSRNAATGLDRLEGGQPSLEQAQLANLADVAYNNHDIQDGLR
SGLTLEQLEGVPLWARHREAEAPFPGVGRMINETIRRIINALIVDLIAGTRAAI
TGAPQSIDDVRAAPPLVAIPESEMRDEARVLKRFELDNLYRHYLVLMRAAKARLIED
LFRVLEDPRLLPYQARFADQPRWIAHYIAGMTDRYAIKEHRRIFAEVL"
complement(1298..2404)
/gene="arob"
/note="RSC2969; RS01328"
complement(1298..2404)
/gene="arob"

```



Db 190831 -----CGCGCAATGCCAATTCGCGCGCGCGCTGCAGTATCGG 190793

QY 56 SerIleLys-----GlyTyrGluProGlnAsp 64

Db 190792 GACGTGGAGCGCATCTGCCAGCTGCGCGTGGCGGATTCACACTATCTCCCGTGACC 190733

QY 65 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 84

Db 190732 CTGCGCGCGGACGCGCTGTGTGACGGCGCACGAAGTGTACCGATCGGCAATCCGCTC 190673

QY 85 SerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly 104

Db 190672 GGGCTGAGCTACCATCAGCAGGAGCTGATCTGCTGCTGCGCACCGACGATGACGGC 190613

QY 105 ArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly 124

Db 190612 CGCTGAG-----AGCGTGCAGACCTCGCGCGCATCTCGCGG 190574

QY 125 GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeu 141

Db 190573 GGCTCAGCGCGCGGCGCTGTTCGATGCCAAGCGCGCGCTGATCGGCATC 190523

RESULT 5

AE007581 11581 bp DNA linear BCT 27-JUL-2001

LOCUS Clostridium acetobutylicum ATCC824 section 69 of 356 of the

DEFINITION complete genome.

ACCESSION AE007581 AE001437

VERSION AE007581.1 GI:15023527

KEYWORDS

SOURCE Clostridium acetobutylicum.

ORGANISM Clostridium acetobutylicum

REFERENCE 1 (bases 1 to 11581)

AUTHORS Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q., Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitt,J., Wolf,Y.I., Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P., Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.

TITLE Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum

JOURNAL J. Bacteriol. 183 (16), 4823-4838 (2001)

MEDLINE 21359325

PUBMED 11466286

REFERENCE 2 (bases 1 to 11581)

AUTHORS Childress,D., Zeng,Q. and Smith,D.R.

TITLE Direct Submission

JOURNAL Submitted (24-JUL-2001) GTC Sequencing Center Production, Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100 Beaver Street, Waltham, MA 02453-8443, USA

FEATURES

Location/Qualifiers

1. 11581

/organism="Clostridium acetobutylicum"

/strain="ATCC 824"

/db\_xref="ATCC:824"

/db\_xref="taxon:1488"

154..819

/gene="CAC0653"

154..819

/gene="CAC0653"

/codon\_start=1

/transl\_table=11

/product="Response regulator (Chev-like receiver domain and HTH DNA binding domain)"

/protein\_id="AAK78630.1"

/db\_xref="GI:15023528"

/translation="MHLIVEDDELSILLKGLKQYNCTCDISHDGEDGLYNLEINS YDAVIDNLPMKGITVCRMRKGISTPVLMTARTDDRVGLGSDGADLYLGP FEFKRLARLHARLNKPSNEIKDLNIDTRAKSVKGEKILITLITAREYDILEL LCYNPNIVASAEIIEHVHWDNDNQSINVRVHTANLRKIKCSGGETLITLKGKY RLC"

813..2354

/gene="CAC0654"

813..2354

/gene="CAC0654"

/codon\_start=1

/transl\_table=11

/product="Sensory transduction histidine kinase"

/protein\_id="AAK78631.1"

/db\_xref="GI:15023529"

/translation="MLKIKSRITALLYSMLTIALIIFIISFCLILNIYINKRPIILDSI TKVGRKSEETNSNNRNIHLTKNFIVEAGGKIKKSNIRVDNSKKDYIIGFVYKKG GVIDYKTEGRYQNKYQNYLETIFNNGKVLIIITLDDYSEDNLSCYGTGNLKYI SKGQIFIDLLKNNSGKLSIVTDKNDSVIRLOKIDDSYDKMFGLSNKGFKMTSG EFEAAMHVLYRRFYIIFIITMIINFALSTYVIRFALAPLLEFFHKKIEQSNAE EIKFVEVPKVNDIYDLTCAYNAMEKIKNSYDDDLQILNSYASHLRLSLAILKAKLQ LGDKINEAYIDKITNTDINAMSTPKLFNDENVDLALVCAKVVEYIEVFRNIEL KLPDNGVELIKGKELWLRPCVANLIDNALKFVDKNKAYNEIRVEVLEDEKDSIRVYD NGIDSESKIDKIFEPYGNMRLSTGIGLAVVKHIMDLHKGRLVLSKKEEYSEISL IFSRR"

3450..3653

/gene="CAC0655"

3450..3653

/gene="CAC0655"

/codon\_start=1

/transl\_table=11

/product="Hypothetical protein, CF-26 family"

/protein\_id="AAK78632.1"

/db\_xref="GI:15023530"

/translation="MKLVNPGRDLTEATEAYSCHVCSSASYKAQSRGFDEVYDSC ECDNWEYFNLYEVSTYIGRKNN"

/complement(3884..4453)

/gene="CAC0656"

/complement(3884..4453)

/gene="CAC0656"

/codon\_start=1

/transl\_table=11

/product="Predicted transposase (3' fragment)"

/protein\_id="AAK78633.1"

/db\_xref="GI:15023531"

/translation="MELESNNOIPYKPYQYKWKKRSPKNDISRRNMEYDAKTDY TCHNRSKSPCLIIHKKSAGSYSEVTVEICEIDNCTLPKCTKAKNKRKMOVSKAF IKKRETSYNNIKTELGTKLRMNSIOVEGAGVLKSDYEPKRELTGRKNSVKTEFILL CPGYNKHLKSKIQNKRTQKYLHELKSLIS"

/complement(4489..5493)

/gene="CAC0657"

/complement(4489..5493)

/gene="CAC0657"

/codon\_start=1

/transl\_table=11

/product="Predicted transposase (5' fragment)"

/protein\_id="AAK78634.1"

/db\_xref="GI:15023532"

/translation="MLVTKLHTKNYNQFNDNLQILPLINLENLIPEDDSVRLLSYLLE GLNKKLYKAYSILGRKSAPKPMFTIISVAYSONIYSSRKIEKACKRQINFKWLQ SFKYPDHATISRFRRKKYLSNEVIEDLYQQVNYLAKKEKLLFENFIDGTRKEANANR YTFNKKVIYKNECKMFDKIIALVEDINLENLAKGFTTGKFTLIDINKLLOWLFEKE KRNFTEVHGIGKRTKIOKWEIQLSQYKROEKYNLKSIFSKRNSYSKTDNDATFPH MKDDHMRNQLKPAYNVQIAVDSEYVTGVGVFDDRNDIATLIPMITNMOKKLVIHTPM "

6172..7728

/gene="CAC0658"

6172..7728

/gene="CAC0658"

/codon\_start=1

/transl\_table=11

/product="Fe-S oxidoreductase"

/protein\_id="AAK78635.1"

/db\_xref="GI:15023533"

/translation="MSTSEQCVEIDKYIDYFKKNTSQPNILAKLFTKTDSSYYVYDTG TNKVLTCGEYEVSLDKINDKDFEISINDADKSGFYEALENVQAIERKEDILRGG KEMFVSGHFEELDKLNYLEDQITLLETEKLCRCRYCIVNESYKRNFGNKDMP TEIAEKADYANLHSGKEKGVGYFGGELVNVNFKLHSHIYSRKIKDKRLTFSLT SNLTMTPEIATYLASIENLILCSIDGPEETHDSYRKDINGNGTFKRAIGLKYLIE AFGDKANKAMVMNVFTPPYSIEKVNKTETFINSLDNLWLPKGLRISKTSYPSGSDVDT PKGCEKNYGMELIQWSESKYSELKSDGKIDMFNEEFVVHPIILRIHGRISAKSESGY



```

PVNGCCIPGSRKLYIVVDGNFQICERINGSPIIGNVFNGIDENKIKNEVIDEFINKSI
KDCSRCAARLNCVSYCHYTDGKLDMEKKRYCIKRSRERALSFLVYFKCAEINPEK
LYLNKIEVS"
/gene=".9084
7765..9084
/gene="CAC0659"
CDS
7765..9084
/gene="CAC0659"
/codon_start=1
/transl_table=11
/product="Predicted Zn-dependent peptidase"
/protein_id="AAK78636.1"
/db_xref="GI:15023534"
/translation="MNKEDEGKNINIKKNTLKNKFAYIICNNKSKGVKLALIVKAGS
LMQGPNGTAHFEHCLDSNVAYILKNVGYEDDNNILQGYTFNFQTVYFIDTIDN
IDEGFVNKILSDICIDRLMEKVKDLIVEIDYRMKCANFGLQNNILPVIIGDKNI
KDFPLGKIKIEKMEFQSVKSPQNSAICICGDIANSIDENKISKYFSFSEFK
NDNEKIEKSTNCPYGEKVVNNVNNKVEKVAQFYLLKDKYKLSKSDTIEVFLSEFK
IKNYILDLTKDSFGVIEFYSKVELHNLKLVLEFKSQSEFVEFLFIDSINKIOAK
GINEDFVKYKRIFYKDLKYSKYEKYYQSEELIKECINFLYDELMFLNSYEACN
LIVDNTKFEKFNLIISQELNNKLNLIIVNGOEKFNLEYMNKFSRFLS"
9167..9370
/gene="CAC0660"
9167..9370
/gene="CAC0660"
/codon_start=1
/transl_table=11
/product="Hypothetical protein, CF-26 family"
/protein_id="AAK78637.1"
/db_xref="GI:15023535"
/translation="MKLVNPGDLTELEATEVSHCHICSSGSGSYTAKMRGYHGGYDAC
ECDDVHANNDSYADTFGIRKNN"
9575..11458
/gene="CAC0661"
9575..11458
/gene="CAC0661"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein (multidrug
resistance protein)"
/protein_id="AAK78638.1"
/db_xref="GI:15023536"
/translation="MKVSNINKKLSGLSSNLIKTKVLLKWKSSLSFATLIITVL
NGIIVPINVSIKYLDSVVNALKNKNSDLNKKVFLWLVFEGVIAFSVIGRVNTP
FSTIQSKLNSISELLIKANDELDSYFENDEKFKIEKANSQAVGTIIQVINGLTO
IVKNTFLGVAIVFQPSIILLCLIMTSIPMFIINIKYKQYINYSKRIKRLA
HYLQVMMDYIPVKEIKRIGDWLEKILISYKDNLDQDKIEKQOLLIMSDVIFN
TIISYGVKIVIPITLQRLSIGSMNYSALTNLEDVSKNTLDNLFALLYSNVLTEN
LFVLDLKLPLNNSKKKEFINEIDSEIFERNVSPKYNPKYALKNVSEKTEANON
CAIVLNGCGKTKLILKRLPYDTEGEIYIDNINKEFNIESLYKGIVFOFPMKY
PFVKKRNGNIDNLDNFELIKTAAKANAYTFIOGLSNKFDTKLOKMWNSGVDSL
GQWQKLAISRAFNSACLILDEPTASIDAKSEYELFKDFKELMGEGTSLISHRST
VRMADRILVLKDGIEVQGNHDSLIKSDGLYAEILYNNQADAYLVNEDKNIDVSKDLK
DIV"
BASE COUNT 4472 a 1147 c 1925 g 4037 t
ORIGIN

Alignment Scores:
Pred. No.: 13.5 Length: 11581
Score: 93.00 Matches: 37
Percent Similarity: 44.00% Conservative: 29
Best Local Similarity: 24.67% Mismatches: 56
Query Match: 9.27% Indels: 28
DB: 1 Gaps: 7

US-10-008-355-2_COPY_522_712 (1-191) x AE007581 (1-11581)
Qy 48 PheThrMetArgMetSerTyrGlySerIlelysglyTyrGluProGlnAspGlyAlaTrp 67
Db 1098 TTTCTATATAGAAAAAGGTGGAGTAATTGATGATATATAAAACCGAAAGGAAGATAT 1157
Qy 68 TyrAsnTyrHisThr-----ThrGlyLysGlyValLeu 78
Db 1158 CAAATAAATACTCAAAATATTTAGAAACTATATTTAATAATGTAAGTATATAATA 1217

```

```

Qy 79 GluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPhe 98
Db 1218 TCTACTATACCTCCA-----GATGATTATTCa--GAGGATAATCTCAGTAAGTGCTAC 1268
Qy 99 ArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSer 118
Db 1269 GGTACTAAAAATTTA---AAGTATTATTTCTAAAGCCAGATTTTTCATTAAGGATTACTT 1325
Qy 119 AsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeu 138
Db 1326 AAAATAATATGAGTGGT-----AAACTATCTATAGTAACAGATAGACATGAGCGTA 1379
Qy 139 IleGlyLeu-----AlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIle----- 154
Db 1380 ATAAGAATCCAGAAAATTTGATGATAGTGTTTATGATAAAATGTTTGGTGAACATTTCAAT 1439
Qy 155 -----GluPheGluProAspLeuGlnArgThrIleSer 165
Db 1440 AATAAGGATTTAAATAACATCAGGAGAAATTTGAAGTCGCGTAATGTCATGTACTCTAT 1499
Qy 166 ValAspIleArgTyrValLeuPheMetile 175
Db 1500 AGGAGATTTATTATATATATATATTATTAATT 1529
RESULT 6
AFAL12H2/c
LOCUS
DEFINITION
Aspergillus fumigatus BAC AFAL12F2.
ACCESSION
AL807577
VERSION
AL807577.1 GI:21627806
KEYWORDS
3-hydroxyisobutyrate dehydrogenase; FAD synthetase;
isopentenyl-diphosphate delta-isomerase; synaptobrevin
ATP-dependent RNA helicase.
SOURCE
Aspergillus fumigatus.
ORGANISM
Aspergillus fumigatus
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE
1 (bases 1 to 79718)
AUTHORS
Harris, D.E., O'Neill, S., Knowles, D.G., Hall, N., Quail, M.,
Woodward, J.R., Denning, D.W., Anderson, M.J. and Barrell, B.
TITLE
Direct Submission
JOURNAL
Submitted (24-JUN-2002) The Sanger Centre, Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, UK
FEATURES
Location/Qualifiers
1..79718
/organism="Aspergillus fumigatus"
/strain="Af293"
/db_xref="taxon:5085"
join(2..1327,1382..4864)
/gene="AFAL12H2.01"
join(<2..1327,1382..4864)
/gene="AFAL12H2.01"
/note="Similar to Saccharomyces cerevisiae hypothetical
269.9 kDa protein in fkh1-met18 intergenic region yll129C
SW:YIM9_YEAST (P40468) (2376 aa) fasta scores: E():
2.1e-67, 32.42% id in 1496 aa
Signal peptide predicted for yll129C by SignalP 2.0 HMM
(Signal peptide probability 0.606, signal anchor
probability 0.000) with cleavage site probability 0.431
between residues 18 and 19"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD37139.1"
/db_xref="GI:21627807"
/translation="IESHGLFFLCSSRRVRFAITVRLITFEDRALGKENTRIIRV
LEADAHEILNVNDEQUSVAERSIRQKRKSASHNTLIECSSEVSDTWSKVFPN
IIRISFETCPFAVTLGRIYCARLVQMKHTITALDSPHPPPYPPYADPAQRALGRN
MTAEIILQKLYLVWACTLNSVQSQSOLANAOHARKSSKQSOOKIKSSARSL
FAVIFLPSAERSSIRNAIVMALGSINKNLYRTFLESLOYAVTTCNEAKIRICAHUR
SPSSPRENRKTDRLRTVTHVYKLTSHFLREPEVYNDWIVNVLVTYTKDLRLFLSDA
EVQNDWFQRLRFHCGMEELFEGINRTPKDPHWIFESRKSAFSLMEDWCWCGSPNQ
AQIAQREENRKRKFAMAHQSESGEGRNTAAAMEIEKKNLRAAALUSAMSLCAGPISITT

```

ESSVQLQFDVGRMLSWIDIIIFNTLSKWHAIIGRRALKNLIIHNKEHSYLLERATEMCH  
VSRPKALESYFVVSTQKTDYPLGFWILGAVLVTLGNOKRETRMKSAXLLRIL  
EERQKSSRLQDDFDTSKDTAVYKLAOFETSRRLAKOHSDLAFTLFSFSLHFRNV  
GPSORNNVAAILPMTVOTMELOIDPNGGPTAKSYMLANLEIFITIRGCNILNPREVQAL  
WQALATGPHGNGVOLVDIFIISICLERKEQNFVEYAKQVVVFLSGPAGSKVIEFFLM  
GVPKNMVQERKDIPTAPDVKSLPPYVADLGTLPVGNKAGLSLGOVALIFVLDMV  
ADFTVLEAVKLVHVLVILMDHYMTLVEQAREMLVHLIHELIAAKIDDDAPAAATRO  
GIEDVFIESRSPQKVVMEYEDNNKEDGDGRVPLSMSTVTRQVNVFESFAVEGVS  
DLWAKEALNATSCPVRLACRSQFVPCISTSLNPRMLADMLARSLNTADREADVQ  
TFSEMLITLTKIIISLAPADLLRYPLQFWITCACLTIIHETEFIESIGMLEKFLDSY  
DMSDPAVVTELIGQCPKWGGEGDGLQNLVYKGLSSESLNRTLDVLRHUSGLPNAL  
IGNSDRLEFLLANMAHFLHOFDPDVPDKTLARATLLARLVAAEGECPKMLASLLGLA  
NGYTKENDLNHIIAEMRSYFFPRQDVQSLIFLMGLTTLTWKFRKYIKRILCVLIP  
EIDMRREVTSHPGDLISPLLRLLQELCPQALEVMDHIMTVSGNPMERHIIHMSMAS  
SSSRAIRKEVEKIQSLYIPEPTGMSIPMPATQSSVTRHNVAHFVTCAEVDPMEOV  
ETVTPVEFHADYNDSEFFPMRADTKMSMDTQMDNGMDIVQKLSLDLDDFEFETEPIN  
SGLDVMDASLRETGYAYDTSAHLVDQQTAPILRKSLARTASTSSPHNGLASRPPN  
QRDSVMSFNTLMPQVSGHTRPTTHARSVTSPANNLYPPSSSNVQFSTPLPGFSE  
SAYLSDDFEEDGLYEFDERLTNNKPTTFVHNSQGRSVTDGSSLESIRSGMRRLTGG  
AAANRERQRDILIRAHRAAVQTANSPRVKVPPPEFLTGTSPNTSPGQ"  
/join(5738, .5906, 6086, .6276, 6472, .7209)  
/gene="Afa12H2.02"  
/join(5738, .5906, 6086, .6276, 6472, .7209)  
/gene="Afa12H2.02"  
/note="Similar to Neurospora crassa putative  
3-hydroxyisobutyrate dehydrogenase g698.5 TR:09C1L3  
(EMBL:AF309689) (338 aa) fasta scores: E(): 1.3e-65,  
56.92% id in 332 aa  
prosite profile:PS00895;3-hydroxyisobutyrate dehydrogenase  
signature Confirmed by InterPro eMOTIF pattern  
match.;8e-5;codon 50-63"  
/codon\_start=1  
/product="putative 3-hydroxyisobutyrate dehydrogenase"  
/protein\_id="CAD37140.1"  
/db\_xref="GI:21627808"  
/translation="MSLTPTVIRIGFVSPKSLSSVLPQLRQSQARTFTTQLR  
DAYWGTGLGOMGLYMAKNLRKIPATDTLLVRDVENATTRFEETKEAKSGGAD  
TKMKVEYASAREAQSTVMTISLPEPEHVKDFVHSILKGLDLPALAEELITDTST  
IDPASKEIANASTHTGFVDAPMSGVGARAGTLSPFMFGASSOGGLIERVRVAV  
LMLGKKAHLPGGAGVSGKLANNYLAINNTATFAMNLGTRWGLDPLKSLADMLNT  
STGRCPMENVNPPGVVETAPASRGVEGFGVGLKKDLRLAITAKESGTPALAE  
TAREVNSVQERGRKDFSVYKWLQDRSQ"  
/join(9141, .9470, 9608, .9697, 9745, .9839, 9901, .10219,  
10253, .10567)  
/gene="Afa12H2.03"  
/join(9141, .9470, 9608, .9697, 9745, .9839, 9901, .10219,  
10253, .10567)  
/gene="Afa12H2.03"  
/note="Similar to Schizosaccharomyces pombe hypothetical  
62.9 kDa protein c29a10.02 in chromosome ii spbc29a10.02  
or spbc365.18 TR:094381 (EMBL:AL034463) (567 aa) fasta  
scores: E(): 3.1e-05, 33.85% id in 319 aa  
pfam:PF00076:RNA recognition motif. (a.k.a. RRM, RBD, or  
RNP domain);0.16;codon 51-119  
prosite pattern:PS50102:Eukaryotic RNA Recognition Motif  
(RRM) profile.;10.326;codon 126-203"  
/codon\_start=1  
/product="conserved hypothetical protein"  
/protein\_id="CAD37141.1"  
/db\_xref="GI:21627809"  
/translation="MKTNPNGMYLQDLESITQEQEPTIPRAVPAMWTPSEVTLAKLCLEN  
REGITNVYIRGFQPTDDEMLHSAARFGKIDRCIAIVDLDTGLCKGYKVAASMSAC  
QLISLPDSKSRSLKDLKDTSTNYICTNVTWTWADLRHHFEYHVYVSEKISRRA  
KTGVSGVGFARDTRIEAKVLGFEHNISKNGVKLLRFDADKAOKMLKQOQNERRA  
VRAGEVNSYFVQGSTPSPSLORLOQTANHLSPNSQVSPSPVGVSGSTWTPATSI  
SYRNSIDIRVQLMKNKPNPSSWSTRNSPATWETTPVYHGLYRGHAIIPNASSAS  
SKTMPESTPCVKARSOGSSPKENMDVDCLSPLSPCKEALMTSPRSAL"  
complement(join(11060, .11087, 11156, .11243, 11312, .12053,  
12210, .12266))  
/gene="Afa12H2.04c"  
complement(join(11060, .11087, 11156, .11243, 11312, .12053,  
12210, .12266))  
/gene="Afa12H2.04c"

/note="Similar to Schizosaccharomyces pombe probable fad  
synthetase spcc1235.04C SW:FAD1\_SCHPO (074841) (265 aa)  
fasta scores: E(): 7e-31, 40.61% id in 261 aa  
pfam:PF01507:phosphoadenosine phosphosulfate reductase  
family;1.8e-11;codon 62-249"  
/codon\_start=1  
/product="probable FAD synthetase"  
/protein\_id="CAD37142.1"  
/db\_xref="GI:21627810"  
/translation="MRESASEPSRALSTSGAARVQSFLOESHPDLSLLAAVQOQOTRI  
SLNVATALSRYKLSLSYNGKDCVLIIILASLHPPHPPEEGGLAYITAIYAL  
PPDSFAVEEFVQMSRAYLAIIVRTTEPTPTTLKSCFEHYSLNPSIKAIIVGTRR  
TDPGHALHFTEDTSGWDFMRIPVIMHWYAEIWAIFIRHLGLIKYCSLYDRGVTSIG  
TSDTHNPKLREHSDAGQGVADGTQSQHYRPAVELTQDEERLGRILNPNFSIQ  
SSSITSFYHSFAPQTPDVGRCSNATELD"  
complement(join(12955, .13492, 13553, .13589, 13671, .13737))  
/gene="Afa12H2.05c"  
complement(join(12955, .13492, 13553, .13589, 13671, .13737))  
/gene="Afa12H2.05c"  
/note="Similar to Schizosaccharomyces pombe DNA j domain  
containing protein spbc1734.05C TR:074746 (EMBL:AL031856)  
(209 aa) fasta scores: E(): 2.1e-17, 41.62% id in 185 aa  
prosite pattern:PS50076;dnaj domain profile.;11.351;codon  
34-98  
smart:SM00271;DnaJ molecular chaperone homology  
domain;1.1e-08;codon 33-90  
pfam:PF00226;DnaJ domain;2.9e-10;codon 34-98"  
/codon\_start=1  
/product="hypothetical DNA j domain containing protein"  
/protein\_id="CAD37143.1"  
/db\_xref="GI:21627811"  
/translation="MSDEQDADALEKEASDFIKVCFISYIILFVAVLDLQPG  
VPESDKLOYRKSLIHPDKTPAAPADAFDLKKAQTTLLDEKARAYLDECIADAR  
RLILREKHYTVDSPELQTEFEKKEWQKTVQVLLBEEARRRRLKAKLQEGEKKRKE  
EEELAKRKRELEQANREFIGRSWREFQKGKRGDGRKKRKKMKVLG"  
/join(13982, .15673, 15761, .15826, 15869, .16207, 16235, .16311,  
16336, .16414)  
/gene="Afa12H2.06"  
/join(13982, .15673, 15761, .15826, 15869, .16207, 16235, .16311,  
16336, .16414)  
/gene="Afa12H2.06"  
/note="Similar to Schizosaccharomyces pombe probable  
ATP-dependent RNA helicase spac1093.05 TR:Q9UFP9  
(EMBL:AL132839) (735 aa) fasta scores: E(): 1.5e-118,  
53.67% id in 749 aa  
smart:SM00490;helicase superfamily C-terminal  
domain;8.5e-25;codon 305-388  
smart:SM00487;DEAD-like helicases superfamily, catalytic  
domain;5.6e-53;codon 66-269  
pfam:PF00271;Helicase conserved C-terminal  
domain;3.6e-27;codon 305-388  
prosite profile:PS00039;DEAD-box subfamily ATP-dependent  
helicases signature Confirmed by InterPro eMOTIF pattern  
match.;8e-5;codon 198-206"  
/codon\_start=1  
/product="probable ATP-dependent RNA helicase"  
/protein\_id="CAD37144.1"  
/db\_xref="GI:21627812"  
/translation="MAPAGPRYGHAKPORSKTLKRRKQDESLIQVDELDILKG  
IFKFSDLPLUSEPTASGLASSHYKTLTDIQSRAISHALKGRDILGAARTSGKTLAPL  
VPLENLYRKQAEHDLGALILSPTRLEIAQIFEVLRKIGRYHTFSAGLVIGSKSLK  
EEOERGRMNLIVCTPGMLHLDQALFDFTYNLQMLVLDADRIIDLQPOQVDAITP  
GHLPKERTOLLFSATQTKVSDIARLSLODPEYVAVHETASSATPSKLOQHYVYITPLP  
QKLDILWSFTSRNLKSKTVVFLSSGKQVFEVSEFHLQPGIPLMLHGHKQKGGRLD  
IVTRFSQSKHCVLFTDVAARGLDPAVDWIQDLPEDADTIHVRGRTARVERGR  
AVLFLDPSEEGMLKRLLEQKKVPIEKINANKQOSTKQDQNMKRPDPEKLYGQRA  
FISYKSVYIQDKDEIKFLKELKDLDEFASLGLPGAPRTIKTGGDTQKRNAPRAA  
HLISDDDDTDEEDGEKKKKKQVTRKYDRMFRNRNQDLAEHYSKLINDQGTWDL  
GSSGDEDDSEKKNKKNVRREKLLKSKLLKFKGKTKLVYDDKGNPHNEYLEED  
EEOFKAGDAKQOAKFLAEAEERLADMEDKEIAKQKREKKERKARERELAEF  
ESETLVQLPPYEGDQDEAPRSKPKVKTETANDREAEPCDQAHTPRQITLED  
LESATGLIG"  
/join(17274, .17294, 17367, .17626, 17679, .17880, 17933, .17974,

gene

18033..18181,18245..18709,18759..19248)  
/gene="Afa12H2.07"  
Join(17274..17294,17367..17626,17679..17880,17933..17974,  
18033..18181,18245..18709,18759..19248)  
/gene="Afa12H2.07"  
/note="Similar to Yersinia pestis putative lysine-specific  
permease lypP TR:CAC90139 (EMBL:AJ414147) (503 aa) fasta  
scores: E(1): 1.8e-78, 42.41% id in 488 aa

## CDS

Alignment Scores:  
Pred. No.: 204 Length: 79718  
Score: 91.50 Matches: 44  
Percent Similarity: 36.84% Conservative: 26  
Best Local Similarity: 23.16% Mismatches: 69  
Query Match: 9.12% Indels: 51  
DB: 8 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x Afa12H2 (1-79718)

Qy 18 AsnAlaTyrAla-----IleGluLysGlyLysArgLeu 28  
|||||  
Db 75473 AACCTCATGCACTCTCAGCTACATCGTCGTGGCGACCGCGCAAGCGCGTTCGCGATT 75414  
Qy 29 PhePheAlaGlyLeuArgGluMet-----TyrProGlyArgAla 41  
|||||  
Db 75413 CCATTGCTTCTCCTCGAGTGAAGCGCAAGTCTCTATCGACCTACCCC----- 75363  
Qy 42 LeuProSerAspAlaAsnPheMetArgMetSerTyrGlySerIleLysGlyTyrGlu 61  
|||||  
Db 75362 ---CCTTCGACGACGATTTCTCTCTCTCCCTGCATACGCGTCCGCGCTTCAACTCG 75306  
Qy 62 ProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGln 81  
|||||  
Db 75305 GAGCTTCTCTCTCTCCAGACATACAAATACCGCC----- 75270  
Qy 82 AspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg----- 99  
|||||  
Db 75269 ---CGCGCTCGGACTCGTTCCTTCAGCCGCGGAGATCGACGCGTCCGCGACATT 75213  
Qy 100 ---ThrLysAsnTyrGlyArgTyrAlaGluAsnGlnLeuHisIleAlaPheLeuSer 118  
|||||  
Db 75212 ATGACAGAGATATTGACGGGTACTGGAGCGTGTGAGCGAATGATTGTTGGTCCAT 75153  
Qy 119 AsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeu 138  
|||||  
Db 75152 AAGACTGATCGCTCGCGCGGAGCGCG-----CACGACTCCGAGTCCGCTACT 75105  
Qy 139 IleGlyLeuAlaPheAspGlyAsnTrpAlaMetSerGlyAspIleGluPheGluPro 158  
|||||  
Db 75104 CGCGTCTCGCGCGCGGATGTGGTGAAGAAATATCAAG----- 75066  
Qy 159 AspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMet----- 174  
|||||  
Db 75065 ---CTGATGGTCTCTCGCTGTCTGTGCTCTCTCTCATCTATCTTTTCGTCGGTATG 75009  
Qy 175 -----IleAspLysTrpGlyClnCys 181  
|||||  
Db 75008 GGATCGCGCTTGCACGCTTGGGGAAGTGT 74979

## RESULT 7

CNS01RG9/c CNS01RG9 180688 bp DNA linear PRI 26-JUN-2001  
LOCUS Human chromosome 14 DNA sequence BAC R-63812 of library RPCI-11  
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.

## ACCESSION

AL157871

## VERSION

AL157871.5 GI:14572586

## KEYWORDS

HTG; HTGS\_ACTIVEFIN.

## SOURCE

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 180688)

AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,  
Levy,M., Eckenberg,R., Bruls,T., Deberardinis,V., Cruaud,C.,  
Gyapay,G., Saurin,W. and Weissbach,J.  
Sequencing of the human chromosome 14  
Unpublished  
2 (bases 1 to 180688)  
Genoscope.  
Direct Submission  
Submitted (26-JUN-2001) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
On Jun 27, 2001 this sequence version replaced gi:11611154.  
----- Genome Center  
Center: Genoscope / Centre National de Sequencage  
Center code: GS  
Web site: http://www.genoscope.cns.fr/  
Contact: SeqRef@genoscope.cns.fr  
-----

The following BAC sequence is oriented from the T7 to the SP6 end.  
Upstream BAC (overlapping the T7 end) : R-775G15  
Downstream BAC (overlapping the SP6 end) : R-362L22 (AC-AL135838)  
----- Summary Statistics  
Assembly program: Phrap: version 2.0  
Quality coverage: 6.78x in Q20 bases; sum-of-contigs  
-----

## Overall quality chart :

Range : bases

0 : bases

1 - 9 : 6

10 - 19 : 83

20 - 29 : 273

30 - 39 : 1283

40 - 49 : 5543

50 - 59 : 10660

60 - 69 : 11794

70 - 79 : 22509

80 - 89 : 49844

90 - 99 : 78693

-----

Percentage of bases with a quality value >= 40 : 99 %

## FEATURES

## source

1..180688  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="R-63812"  
/clone\_lib="RPCI-11"  
37623..37869  
/note="matching EMBL:267128  
RHdb:RH53816  
RHdb:RH3709  
dbSTS:STS5886  
Identified using the e-PCR software (G. Schuler)"  
74837..74962  
/note="matching EMBL:G11457  
RHdb:RH1924  
dbSTS:STS19283  
Identified using the e-PCR software (G. Schuler)"  
85576..85699  
/note="matching EMBL:M77698  
RHdb:RH44091  
RHdb:RH1618  
dbSTS:STS16477  
Identified using the e-PCR software (G. Schuler)"  
86726..86933  
/note="matching EMBL:H38818  
RHdb:RH68412  
dbSTS:STS48319  
Identified using the e-PCR software (G. Schuler)"  
87142..87270  
/note="matching EMBL:T78109  
RHdb:RH25853

```
dbSTS:STS24160
Identified using the e-PCR software (G. Schuler)"
STS
90436..90668
/notes="matching EMBL:D11738
RHdb:RH53906
dbSTS:STS42918
Identified using the e-PCR software (G. Schuler)"
STS
99113..99287
/notes="matching EMBL:H37767
RHdb:RH77775
RHdb:RH15727
dbSTS:STS13361
Identified using the e-PCR software (G. Schuler)"
STS
141786..141911
/notes="matching EMBL:H28114
RHdb:RH53813
dbSTS:STS26412
Identified using the e-PCR software (G. Schuler)"
STS
142470..142596
/notes="matching EMBL:T63858
RHdb:RH53514
dbSTS:STS32482
Identified using the e-PCR software (G. Schuler)"
STS
142501..142639
/notes="matching EMBL:X59892
RHdb:RH70848
dbSTS:STS50705
Identified using the e-PCR software (G. Schuler)"
STS
144853..145036
/notes="matching EMBL:AA085408
RHdb:RH63341
dbSTS:STS45273
Identified using the e-PCR software (G. Schuler)"
STS
150077..150193
/notes="matching EMBL:H13689
RHdb:RH44843
dbSTS:STS37911
Identified using the e-PCR software (G. Schuler)"
STS
173770..173952
/notes="matching EMBL:AA251131
RHdb:RH103649
dbSTS:STS71178
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 45075 a 44223 c 46191 g 45199 t
ORIGIN

Alignment Scores:
Pred. No.: 550 Length: 180688
Score: 91.50 Matches: 42
Percent Similarity: 32.22% Conservative: 16
Best Local Similarity: 23.33% Mismatches: 57
Query Match: 9.12% Indels: 65
DB: 9 Gaps: 7

US-10-008-355-2_COPY_522_712 (1-191) x CNS01RG9 (1-180688)
Qy 28 LeuPheAlaGlyLeuArgGluMetTyPrProGlyArgAlaLeuProSerAspAlaAsn 47
Db 95029 CTTGTGACTGTAGGATGGCTCAGCCCTATGCACACCATCCCTTCATTCTCCCAACAGC 94970
Qy 48 PheThrMetArgMet-----SerTyRGlySerIleLysGlyTyRGlU 61
Db 94969 CATACAGGTGCTGCTGGCACCCTATTTTGCAGTCAGGGTACCAAGTTGGGAGGGTC 94910
Qy 62 ProGlnAspGlyAlaTyPrAsnTyRHisThrGlyLysGlyValLeuGluLysGln 81
Db 94909 CCTGTCCAAGGC-----CACACGCTAGTTGCTGTGAGCCAGCCGCGAG 94865
Qy 82 AspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLys 101
Db 94864 GAGCCAGTCAGCCCATCTTTGGCTTCAGCAG-----GATCAGCTGGGGCCAGGC 94814
Qy 102 AsnTyRGly-----Arg 105
```

```
Db 94813 ACCCTGGAGCGCAATGCAAGAACACAGAGGGGCTCGCCAGGCGCAAGGTCCAGGTCC 94754
Qy 106 TyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGly 125
Db 94753 CATGCTCCTGGGTCTCAGCTTCCTTGTCTTACACACCCCTGACCTCATG----- 94700
Qy 126 AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla----- 142
Db 94699 -----GTACCCACCTACCCAGGAGGAGGATGATGCTCGGTCCACAGAGCCAC 94649
Qy 143 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 162
Db 94648 TTCCAGGGAACCTGGGAAGG----- 94628
Qy 163 ThrIleSerValAspIleArgTyRValLeuPheMetIleAspLysTrpGlyGlnCysPro 182
Db 94627 -----CGGTGGGCGCAGTGTCCA 94610

RESULT 8
CNS01DVI 187710 bp DNA linear PRI 04-OCT-2001
LOCUS Human chromosome 14 DNA sequence BAC R-362L22 of library RPCI-11
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL135838
VERSION ALL135838.6 GI:15982201
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187710)
AUTHORS Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., DeBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissensbach, J.
TITLE Sequencing of the human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187710)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT On Oct 5, 2001 this sequence version replaced gi:13872725.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS.
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2644121 (AC=AL163974)
Downstream BAC (overlapping the SP6 end) : R-63812 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 6.92x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 : 1
1 - 9 : 53
10 - 19 : 407
20 - 29 : 1172
30 - 39 : 5394
40 - 49 : 11391
50 - 59 : 10988
60 - 69 : 11109
70 - 79 : 20167
80 - 89 : 49419
90 - 99 : 77609
-----
```

Percentage of bases with a quality value >= 40 : 96 %.

```

FEATURES
  source
    Location/Qualifiers
      1..187710
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="14"
        /clone_lib="RPCI-11"
        /clone_lib="RPCI-11"
        /note="matching EMBL:RI7876"
        RHD:RH26378
        dbSTS:STS8590
        Identified using the e-PCR software (G. Schuler)"
        87948..88130
          /note="matching EMBL:AA251131"
          RHD:RH103649
          dbSTS:STS71178
          Identified using the e-PCR software (G. Schuler)"
          111707..111823
            /note="matching EMBL:H13689"
            RHD:RH44843
            dbSTS:STS37911
            Identified using the e-PCR software (G. Schuler)"
            116864..117047
              /note="matching EMBL:AA085408"
              RHD:RH65341
              dbSTS:STS45273
              Identified using the e-PCR software (G. Schuler)"
              119261..119399
                /note="matching EMBL:X59892"
                RHD:RH70848
                dbSTS:STS50705
                Identified using the e-PCR software (G. Schuler)"
                119304..119430
                  /note="matching EMBL:T63858"
                  RHD:RH53514
                  dbSTS:STS32482
                  Identified using the e-PCR software (G. Schuler)"
                  119989..120114
                    /note="matching EMBL:H28114"
                    RHD:RH53813
                    dbSTS:STS26412
                    Identified using the e-PCR software (G. Schuler)"
                    126212..126786
                      /note="matching EMBL:H37767"
                      RHD:RH77775
                      RHD:RH15727
                      dbSTS:STS13361
                      Identified using the e-PCR software (G. Schuler)"
                      171231..171463
                        /note="matching EMBL:D11738"
                        RHD:RH53906
                        dbSTS:STS42918
                        Identified using the e-PCR software (G. Schuler)"
                        174629..174757
                          /note="matching EMBL:T78109"
                          RHD:RH25853
                          Identified using the e-PCR software (G. Schuler)"
                          174966..175173
                            /note="matching EMBL:H38818"
                            RHD:RH68412
                            dbSTS:STS48319
                            Identified using the e-PCR software (G. Schuler)"
                            176200..176323
                              /note="matching EMBL:M77698"
                              RHD:RH44091
                              RHD:RH1618
                              dbSTS:STS16477
                              Identified using the e-PCR software (G. Schuler)"
                              186933..187058
                                /note="matching EMBL:G11457"
                                RHD:RH1924
  dbSTS:STS19283
  Identified using the e-PCR software (G. Schuler)"
  BASE COUNT 48986 a 48210 c 45521 g 44992 t 1'others
  ORIGIN
  Alignment Scores:
  Pred. No.: 576 Length: 187710
  Score: 91.50 Matches: 42
  Percent Similarity: 32.22% Conservative: 16
  Best Local Similarity: 23.33% Mismatches: 57
  Query Match: 9.12% Indels: 65
  DB: 9 Gaps: 7
  US-10-008-355-2_COPY_522_712 (1-191) x CNS01DVI (1-187710)
  Qy 28 LeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsn 47
  Db 166870 CTTGTGACTGTAGGATGGCTACGCCCTATGCACACCATCCCTTCATCTCCCAACAGC 166929
  Qy 48 PheThrMetArgMet-----SertYrGlySerIleLysGlyTyrGlu 61
  Db 166930 CATACAGGTGCGTGTGGCACCCTCATTTTGCAGTGCAGGTACACAGTGGGAGGTC 166989
  Qy 62 ProGlnAspGlyAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuGluLysGln 81
  Db 166990 CCTGTCCAAGGC-----CACACAGCTAGTTGGTGTGAAGCAGCGCCAG 167034
  Qy 82 AspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLys 101
  Db 167035 GAGCCAGTCACCCATCTTTGCCCTTCAGCAG-----GATCAGCTGGGCGCAGC 167085
  Qy 102 AsnTyrGly-----Arg 105
  Db 167086 AACCTGAGCGCAATGCAAGAACAGAGGGGCGCTGGCCAGGCAAGGTCAGAGTCC 167145
  Qy 106 TyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGly 125
  Db 167146 CATGCTCCTGGGTCTCAGCTTCCTTGTCTTCTATACACACCCCTGACCTCATG----- 167199
  Qy 126 AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla----- 142
  Db 167200 -----GTACCCACCTACCCAGGAGAGCAGATGCATGTGGTCCAGAGGCCAC 167250
  Qy 143 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 162
  Db 167251 TTCCAGGGAAGTGGAAAG----- 167271
  Qy 163 ThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysPro 182
  Db 167272 -----CGTGGGCGCAGTGTCCA 167289
  RESULT 9
  AIU52600
  LOCUS
  DEFINITION
  Avian infectious bronchitis virus Q3/88 nucleocapsid protein mRNA,
  complete cds.
  ACCESSION
  U52600
  VERSION
  U52600.1 GI:1853992
  KEYWORDS
  Avian infectious bronchitis virus strain=Q3/88.
  SOURCE
  ORGANISM
  Avian infectious bronchitis virus
  Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
  Coronaviridae; Coronavirus.
  REFERENCE
  1 (bases 1 to 1554)
  Sapats,S.I., Ashton,F., Wright,P.J. and Ignjatovic,J.
  Novel variation in the N protein of avian infectious bronchitis
  virus
  JOURNAL
  Virology 226 (2), 412-417 (1996)
  MEDLINE
  97124667
  PUBMED
  8955062
  REFERENCE
  2 (bases 1 to 1554)
  Sapats,S.I., Ashton,F., Wright,P.J. and Ignjatovic,J.
  Direct Submission

```



```

Db 1696 -----GGTCTTGTTAGCGCATATTGCCAGCAGCAACAGCTGGAGCGCAATGCT 1746
Qy 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 97
Db 1747 TCTGATAAGAGGCGGCAACAGCGCGGAATTTACTGTG-----1785
Qy 98 PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeu 117
Db 1786 -----AATTTTGGCGAGAAAAAATTACCGGCACGTTA-----1818
Qy 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArg 137
Db 1819 -----ACCGCTGAAACACAGCGGAGGAGCAACCTTTTACCATTGATGGTAAAG 1863
Qy 138 LeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGlu 157
Db 1864 ATTGAGGCAACGGTTTTCGGTACG-----GCCAAAACCTGCTGAATAGTTTGTAT 1917
Qy 158 ProAspLeuGlnArgThrIleSerValAspIleArgTyrVal 171
Db 1918 CTCGATCAAAAAAATACCACCGCAGCGCTAAGGCATATATC 1959

RESULT 11
NMTBP2163
LOCUS NMTBP2163 2163 bp DNA linear BCT 31-JAN-1996
DEFINITION N.meningitidis DNA for tbp2 gene (strain BZ163).
ACCESSION Z50731
VERSION 250731.1 GI:1177568
KEYWORDS Tbp2; Transferrin-binding protein 2.
SOURCE Neisseria meningitidis.
ORGANISM Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 2163)
AUTHORS Legrain,M., Findel,A., Villevall,D., Quentin-Millet,M. and
Jacobs,E.
TITLE Molecular characterization of hybrid transferrin-binding protein
2's from Neisseria meningitidis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2163)
AUTHORS Legrain,M.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1995) Legrain M., Transgene, Molecular Genetics,
11, rue de Molsheim, Strasbourg, Bas-Rhin, France, 67000
FEATURES
source
1. 2163
/organism="Neisseria meningitidis"
/strain="BZ163"
/db_xref="taxon:487"
1. 2136
/gene="tbp2"
1. 2136
/gene="tbp2"
/transl_start=1
/transl_table=11
/product="transferrin-binding protein 2 precursor"
/protein_id="CAA90598.1"
/db_xref="GI:1177569"
/db_xref="SPTREMBL:Q51284"
/translotion="MNPPLVNOAAWLVFLLSACLGGGSGFDLSDVDEAPRPAPKY
QQVSEKPAQKDDGGYGFAMRLRRNRHQAEDKVELNPNDEETGLPSKPNLPE
RQSVIDKVTDDGSNTYTSPLYLQSNHQSTNSGAPKNEVDYKNFYVYSGWF
YKAESEREFSKIFKSGDDGYIFYHCKDPSQLPTSEKVIYKGVWHEFTDEKQKF
NDILETSKGGDDGYSGFGDDGETTSNRTDSNLNDKHEGYFTSNLEVDGSKKLTK
LIRNRVNTATNDKYTYQYISLDAQITGNRFGKALATPDGTGGKLHPFVSDSS
LSGFFPGKGEELGFRLESDDKVAVGSAKTDKTENGAVASGTTDAASNGAAGTS
SENGLTVDLDAVELKLDNFDNFVSNAAQLVVDGIMPLPETSESGNANGQ
TNGGTATFRKFDHTPESDKDAQAGTQTNGAQTASNTAGTNGTKTKYSEVCCSNLN
YLYKGMLTRKNSQAGESSQDAKTEQVGOSMFQLOGERTDEKTPSPONTVYRG
SMYGHASSTWSGNSGADKEGNEAETVNFPEKKITGLTAENRQEAFTFLDKIEG
NGFSGTAKTAEGLDQKNTNTPKAVITDKYVGGFYGPKAELGGWFAIQGDQKT
ENTTVASGNGNSASSATVVFGAKKRQKPVQ"

```

```

sig_peptide 1. 60
mat_peptide /gene="tbp2"
61. 2133
/gene="tbp2"
/product="transferrin-binding protein 2"
BASE COUNT 706 a 480 c 529 g 448 t
ORIGIN
Alignment Scores:
Pred. No.: 3.79 Length: 2163
Score: 90.00 Matches: 38
Percent Similarity: 37.93% Conservative: 28
Best Local Similarity: 21.84% Mismatches: 66
Query Match: 8.97% Indels: 42
DB: 1 Gaps: 7
US-10-008-355-2_COPY_522_712 (1-191) x NMTBP2163 (1-2163)
Qy 2 LysSerValIleAlaAlaAargAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 21
Db 1552 AAAAAACAGCAAGTCCGCGATGTCAGGAGGAGAAAGCAGTAGTCAAGCTGATCTAAACG 1611
Qy 22 IleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAla 41
Db 1612 GAACAAGTTGGACAAGATATGTTCTCCAAGC-----GAGCGCACCGCATGAAAAAGAG 1665
Qy 42 LeuProSerAspAlaAsnPhetThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu 61
Db 1666 ATTCCAAGCGGACAAACATCGTTTATCGG-----1695
Qy 62 ProGlnAspGlyAlaTrpTyrAsnTyr-----HisThrThrGlyLysGly---Val 77
Db 1696 -----GGGTCTGTGTACGGCATATTGCCAGCAGCACAAAGCTGGAGCGCAATGCT 1746
Qy 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 97
Db 1747 TCTGATAAGAGGCGGCAACAGCGCGGAATTTACTGTG-----1785
Qy 98 PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeu 117
Db 1786 -----AATTTTGGCGAGAAAAAATTACCGGCACGTTA-----1818
Qy 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArg 137
Db 1819 -----ACCGCTGAAACACAGCGGAGGAGCAACCTTTTACCATTGATGGTAAAG 1863
Qy 138 LeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGlu 157
Db 1864 AVTGAGGCAACGGTTTTCGGGTACG-----GCCAAAACCTGCTGAATAGTTTGTAT 1917
Qy 158 ProAspLeuGlnArgThrIleSerValAspIleArgTyrVal 171
Db 1918 CTCGATCAAAAAAATACCACCGCAGCGCTAAGGCATATATC 1959

```

## RESULT 12

AK023006

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1

Ishigai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,

Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,

Tanabe,T., Nomura,Y., Foglia,S., Komai,F., Hara,R., Takeuchi,K.,

Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,

Wakamatsu,A., Nakamura,Y., Negahari,K., Masuho,Y. and Oshima,A.

TITLE	NEDO human cDNA sequencing project	Db	898	TGCGCAGTGGAGTTCAACTTC	918
JOURNAL	Unpublished	RESULT 13			
REFERENCE	2 (bases 1 to 2794)	BC021506			
AUTHORS	Isogai,T. and Otsuki,T.	LOCUS			
TITLE	Direct Submission	DEFINITION			
JOURNAL	Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)				
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'-3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.	ACCESSION			
		BC021506			
		VERSION			
		BC021506.1			
		KEYWORDS			
		house mouse.			
		SOURCE			
		ORGANISM			
		Mus musculus			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
		1 (bases 1 to 2908)			
		Strausberg,R.			
		Direct Submission			
		Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
		NIH-MGC Project URL: http://mgc.nci.nih.gov			
		Contact: MGC help desk			
		Email: cgaps-r@mail.nih.gov			
		Tissue Procurement: Jeffrey Green M.D.			
		CDNA Library Preparation: Life Technologies, Inc.			
		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
		DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),			
		Gaithersburg, Maryland;			
		Web site: http://www.nisc.nih.gov/			
		Contact: nisc.mgc@hri.nih.gov			
		Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.			
		Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
		Series: IRAC Plate: 55 Row: i Column: 3			
		This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.			
		FEATURES			
		source			
		1..2908			
		/organism="Mus musculus"			
		/db_xref="taxon:10090"			
		/map="FVB/N"			
		/clone="MGC:36621 IMAGE:5347099"			
		/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."			
		/clone_lib="NCI_CGAP_Mam6"			
		/lab_host="DH10B"			
		/note="Vector: pCMV-SPORT6"			
		110..2689			
		/codon_start=1			
		/product="Similar to E1B-55kDa-associated protein 5"			
		/protein_id="AAH21506.1"			
		/db_xref="GI:18204832"			
		/translation="MDVRLKVNLELREQLRGLTRGLKAEALRLAALAEPEPEDERELEADDDPLGPNNEEVEEGSELEGTAQPPGLQHPHPEGSGFPGDGHVMDNITRNQOFYETPVIKQENESSYDRRPLDMEFPQQVYHPELTKEMKQAPSFLLPEASQLKTRDPQFQNRKRPFFENRGYFHEHRRGRSQPPADEDEDDDLVAIDTYNCDLHFVKARDRSSGYPLTIEGFAYLWSGARASYGVRRGRVCFENKIDEEISVKHLPSTCDLPHVVRIGNSLSDSCSTOLGPEPFYSGYGTGKSTNSRFENKIDKFAENDVIGCPAFDECGNDVLSFTKNGKRWGMAFRIKAEALGGQALYPHVLVKNCVAFENVCORAEVYCSVLPGFTHIQLPLSERIRGTIGTPKSKAECIELMMVGLPRAKGTWTKAAASNPSSKYNILGTNAIMDKMRVGLRQRNAGRWDLVLIQATQCLNRLIQIARAKRKNYILDQTNVYGSQRRKRRPFEGFQKAIIVICTPTDEDLKORTVTRTDEEGKDVDPDHALEKMANF			
		CDS			
		1..2794			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone="NT2RP2005168"			
		/cell_line="NT2"			
		/note="teratocarcinoma"			
		/clone_lib="NT2RP2"			
		/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."			
BASE COUNT	687 a 851 c 735 g 521 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	5.96	Length:	2794		
Score:	89.50	Matches:	42		
Percent Similarity:	38.32%	Conservative:	22		
Best Local Similarity:	25.15%	Mismatches:	56		
Query Match:	8.92%	Indels:	47		
DB:	9	Gaps:	8		
US-10-008-355-2_COPY_522_712 (1-191) x AK023006 (1-2794)					
Oy	19	AlaTyrAlaIleGlyLysGlyLysArgPhePheAlaGlyLeuArgGluMetTyrPro	38		
Db	514	AGCTATGGGTGCAGNAGGGCGGTATGCTTCGAGATGAAGATCAATGAGAAATCTCC	573		
Oy	39	GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet-----	52		
Db	574	GTGAAGCACCTTCGCTCTACAGACGCTGACCCACGCTGCTCCGTGCTGCTCCCTG	633		
Oy	53	---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHis	71		
Db	634	GACTCTGCACACCCAGCTAGGCGAAGAGCCTTCTCC-----TATGGCTATGGA	684		
Oy	72	ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln	91		
Db	685	GGCACTGGGAAG-----AAGTCCACCAAT-----	708		
Oy	92	GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly	110		
Db	709	-----AGCGGTGTTGAAAACACTACGAGACAAAGTTGACAG-----	744		
Oy	111	GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly-----	124		
Db	745	-----AACGATGTGATTGCTGCTTTTGGGATTTTGA	777		
Oy	125	---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe	143		
Db	778	TGTGGAATGACGTGGAACCTGCTTTTACCAAGATGAAAGTGGATGGCATTGCTTTC	837		
Oy	144	AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr	163		
Db	838	CGAATCCAGAGAGAACCTTGGGGGCTCAGGCCCTCTATCTCTGCTGCTGAAGAA	897		
Oy	164	IleSerValAspIleArgTyr	170		



TLDPVGDLEVLFIQLQREADKLVRQYNEEGKAGPPPEKFRDNRGGGFRGGGG  
GQYDNRGPPGNGRQFNQNGGGGNGYRGNGRGGGNGYRGNGRGGGNGYRGNGRGGG  
NNSNNGRNYNRAPQPP  
SSANTSTPTVSSYTPQPSYQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP  
PYTPPP  
STQ"

BASE COUNT 724 a 850 c 814 g 520 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.16 Length: 2908  
Score: 89.50 Matches: 42  
Percent Similarity: 38.32% Conservative: 22  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 8.92% Indels: 47  
DB: 10 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BC021506 (1-2908)  
Qy 19 AlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPro 38  
Db 857 AGCTATGGGTCAGAGAGGCGGTGTGTTTCAAAATGAAGATCAATGAGAAATCTCC 916  
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52  
Db 917 GTGAAGCATCTTCCATCTACAGAGCCGACCCACCGCTGTGCTGATTTGGCTGCTCTTG 976  
Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHis 71  
Db 977 GACTCTTGCACCATCAGCTAGGTGAGAGGCTTTTCC-----TATGGTATGGA 1027  
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91  
Db 1028 GGCACAGGGAAG-----AAGTCTACCAAT----- 1051  
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110  
Db 1052 -----AGCGGCTTTGAAACTACGAGACAAATTTGCTGAG----- 1087  
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly----- 124  
Db 1088 -----AAGCATGTGATGCTGCTTTCGCGACTTTGAA 1120  
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuLysGlyLeuAlaPhe 143  
Db 1121 TGTGGAATGATGTGGAGCTTCTTTTACCAAGAATGCAAGTGGATGGCATTGCCTTC 1180  
Qy 144 AspGlyAsnTyrGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163  
Db 1181 CGAATCCAGAAAGAGGCTTATAGGGGGTCAGGCCCTCTATCTCTGCTGCTGGAAGAAT 1240  
Qy 164 IleSerValAspIleArgTyr 170  
Db 1241 TGTGCTGTGGAATTAACCTTT 1261

RESULT 14  
BC027713  
LOCUS  
DEFINITION Homo sapiens, similar to E1B-55kDa-associated protein 5, clone  
ACCESSION BC027713  
VERSION BC027713.1 GI:20379473  
KEYWORDS MGC.  
SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 2956)  
Direct Submission  
Submitted (08-APR-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK  
COMMENT  
NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgaaps-x@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland:  
Web site: http://www.nisc.nih.gov/  
Contact: nisc\_mgc@nigr.nih.gov  
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Laric, P., Legaspi, R., Maduro, O.L.,  
Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J.,  
Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C.,  
Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILLNL at: http://image.llnl.gov  
Series: IRAL Plate: 26 Row: e Column: 4.  
FEATURES  
Location/Qualifiers  
1. .2956  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:16645 IMAGE:4123077"  
/tissue\_type="Muscle, rhabdomyosarcoma"  
/clone\_lib="NIH\_MGC\_17"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
63. .2477  
/codon\_start=1  
/product="Similar to E1B-55kDa-associated protein 5"  
/protein\_id="AAH27713.1"  
/db\_xref="GI:20379474"  
/translation="MDVRLKVNLEELRRLRGLDRLKLAELARLQAALEAEPPD  
ERLADDEPRGPHINEEVETEGSGLETAQPPPGLOPHAPGPGSGDGHYAMD  
NITRQNGYDTQVLIKQENESYBRPLEMEQQQAYREMTENKQGAFTSFLPEASQ  
LKPDQFQSKRPKRYNENRGYFEHREDRRGRSPQPAEDEDDEDDTLVDITYNC  
DLHKVARDSSGYPLTIEGFAYLWSARGSYGVRGKVCFMKINEISVKHLPSFE  
PDHVRIGMSLDSCSTGLGEFFSYGCTGKSTNSRFENYDKFAENDVIGCFAD  
FECNDVETSLFKNGKMGAFRIKQALGQALYPHLVKNCVAFENFGORAPYCS  
VLGFTFTLHPLSERTRGTVPKSKAECELLMMVGLPAAGKTTWATKHAASNPSSKY  
NILGTNAIMDKRMVGLRQRYNAGRWDLVLCQATCGLNRLIQAARKKRYLLDQTN  
VYSGAQRKMRPEGFGFQKALVICPTDEDLKRITKIDEGKDPDPAHLEKANET  
LPDVGDELDELVEITELQREADKLVRQYNEEGKAGPPPEKFRDNRGGGFRGGGG  
GQYDNRGPPGNGRQFNQNGGGGNGYRGNGRGGGNGYRGNGRGGGNGYRGNGRGGG  
NNSNNGRNYNRAPQPP  
SSANTSTPTVSSYTPQPSYQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP  
PYTPPP  
STQ"

BASE COUNT 722 a 880 c 826 g 528 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.28 Length: 2956  
Score: 89.50 Matches: 42  
Percent Similarity: 38.32% Conservative: 22  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 8.92% Indels: 47  
DB: 9 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BC027713 (1-2956)  
Qy 19 AlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPro 38  
Db 807 AGCTATGGGTCAGAGAGGCGCGGTGTGTTTCAAAATGAAGATCAATGAGAAATCTCC 866  
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52  
Db 867 GTGAAGCATCTTCCATCTACAGAGCCGACCCACCGCTGTGCTGCTGCTGCTG 926

```

QY 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHis 71
Db 927 GACTCTGCTGCAGCACCAGCTAGCGAAGAGCTTCTCC-----TATGGCTATGGA 977
QY 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
Db 978 GGCACCTGGGAAG-----AAGTCCACCAAT-----1001
QY 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110
Db 1002 -----ACCGGTTTGAATACTACGACACAGTTTGCAGAG-----1037
QY 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly-----124
Db 1038 -----AAGCATGTGATTGCTGCTTGGCGATTGTTGAA 1070
QY 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 1071 TGTGGAATGACGTGGAACTGCTCTTTTACCAAGAATAAGTGGATGGGCAATTGCTTTC 1130
QY 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 1131 CGAATCCAGAAGAGAGCTTGGGGGCTCAGGCCCTCTATCTCATGCTGCTGGTGAAGAAAT 1190
QY 164 IleSerValAspIleArgTyr 170
Db 1191 TGGCGACGTGGAGTTCAACTTC 1211

RESULT 15
BC009988 3048 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, clone MGC:16706 IMAGE:4127873, mRNA, complete cds.
DEFINITION BC009988
ACCESSION BC009988.1 GI:14603012
VERSION MGC.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3048)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 26 Row: h Column: 8.
Location/Qualifiers
1. .3048
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
Source

```

```

/clone="MGC:16706 IMAGE:4127873"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH_MGC_19"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
CDS
92..409
/codon_start=1
/product="Unknown (protein for MGC:16706)"
/protein_id="AAH09988.1"
/db_xref="GI:14603013"
/translation="MDVRLKVKVNLREELQRRGLDTRGLKAELARLQALAEPEPDD
ERELDADDEGPRGHINEEMPWTILPRTNSTIPKSSNKKTSQATRGDHWKSSSRPI
VOK"
BASE COUNT 761 a 922 c 820 g 545 t
ORIGIN
Alignment Scores:
Pred. No.: 6.52 Length: 3048
Score: 89.50 Matches: 42
Percent Similarity: 38.32% Conservative: 22
Best Local Similarity: 25.15% Mismatches: 56
Query Match: 8.92% Indels: 47
DB: 9 Gaps: 8
US-10-008-355-2_COPY_522_712 (1-191) x BC009988 (1-3048)
QY 19 AlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrPro 38
Db 730 AGCTATGGGTGACAGAGGGCGGTGATGCTTCGAGATGAAGATCAATGAGGAATCTCC 789
QY 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet-----52
Db 790 GTGAAGCACCTTCGCTACAGAGCCCTGACCCACCTGGTCCGTATCGGCTGGTCCCTG 849
QY 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHis 71
Db 850 GACTCTGCGACACCCAGCTAGCGAAGAGCTTCTCC-----TATGGCTATGGA 900
QY 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
Db 901 GGCACCTGGGAAG-----AAGTCCACCAAT-----924
QY 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110
Db 925 -----AGCCGGTTTGAATACTACGACACAGTTCGACAG-----960
QY 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly-----124
Db 961 -----AAGCATGTGATTGCTGCTTGGCGATTGTTGAA 993
QY 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 994 TGTGGAATGACGTGGAACTGCTCTTTTACCAAGAATAAGTGGATGGGCAATTGCTTTC 1053
QY 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 1054 CGAATCCAGAAGAGAGCTTGGGGGCTCAGGCCCTCTATCTCATGCTGCTGGTGAAGAAAT 1113
QY 164 IleSerValAspIleArgTyr 170
Db 1114 TGGCGACGTGGAGTTCAACTTC 1134
Search completed: May 23, 2003, 11:35:49
Job time : 1208.2 secs

```

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 09:42:03 ; Search time 16.2868 Seconds  
(without alignments)  
3596.482 Million cell updates/sec

Title: US-10-008-355-2\_COPY\_522\_712

Perfect score: 1003  
Sequence: 1 SRSVTAARAIAQADANAY.....LFMDKWGQCPRLIQELKLI 191

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US1008355/runat\_16052003\_110403\_2582/app\_query.fasta\_1.1230  
-DB-issued\_Patents\_NA -QMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US1008355.qcgn\_1.1.47\_erunat\_16052003\_110403\_2582 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	977	97.4	1974	4	US-09-221-017B-726
2	427.5	42.6	2384	4	Sequence 726, App
3	89.5	8.9	10862	4	Sequence 1045, Ap
4	83	8.3	1083	4	Sequence 1, Appli
5	81	8.1	849	4	Sequence 1520, Ap
6	81	8.1	849	4	Sequence 3, Appli
7	77.5	7.7	841	4	Sequence 39, Appl
8	77.5	7.7	4403765	4	Sequence 2, Appli
9	77.5	7.7	4411529	4	Sequence 1, Appli
10	74.5	7.4	11672	4	Sequence 2, Appli
11	74.5	7.4	11459	4	Sequence 3, Appli
12	74	7.4	2989	3	Sequence 9, Appli

13	74	7.4	2989	3	US-09-135-166-9	Sequence 9, Appli
14	74	7.4	2989	4	US-08-942-046-9	Sequence 9, Appli
15	73.5	7.3	2934	4	US-09-206-942-52	Sequence 52, Appl
16	73.5	7.3	2952	4	US-09-206-942-50	Sequence 50, Appl
17	73.5	7.3	3270	4	US-08-637-732A-1	Sequence 1, Appli
18	73.5	7.3	10718	3	US-08-325-426B-1	Sequence 1, Appli
19	73	7.3	5741	1	US-07-706-699-4	Sequence 4, Appli
20	73	7.3	5741	1	US-07-998-931-4	Sequence 4, Appli
21	73	7.3	11517	1	US-07-920-281C-1	Sequence 1, Appli
22	73	7.3	11517	4	US-08-466-277-1	Sequence 1, Appli
23	72.5	7.2	842	3	US-08-733-230-3	Sequence 3, Appli
24	72.5	7.2	842	4	US-08-953-326-3	Sequence 3, Appli
25	72.5	7.2	2230	3	US-08-448-194-7	Sequence 7, Appli
26	72.5	7.2	2230	4	US-08-867-921-7	Sequence 7, Appli
27	72	7.2	506	1	US-08-469-802B-7	Sequence 7, Appli
28	72	7.2	506	2	US-08-267-803B-7	Sequence 7, Appli
29	72	7.2	2174	4	US-09-613-444-1	Sequence 1, Appli
30	72	7.2	2974	3	US-08-433-522A-7	Sequence 7, Appli
31	72	7.2	2974	3	US-09-135-166-7	Sequence 7, Appli
32	72	7.2	2974	4	US-08-942-046-7	Sequence 7, Appli
33	72	7.2	10660	2	US-08-267-803B-8	Sequence 8, Appli
34	72	7.2	10660	4	US-09-041-886-16	Sequence 16, Appl
35	71.5	7.1	1168	1	US-08-230-047-4	Sequence 4, Appli
36	71.5	7.1	1197	2	US-08-829-026A-5	Sequence 5, Appli
37	71.5	7.1	11464	4	US-08-991-840A-2	Sequence 2, Appli
38	71	7.1	927	1	US-08-507-431-5	Sequence 5, Appli
39	71	7.1	927	2	US-08-902-655A-5	Sequence 5, Appli
40	71	7.1	927	3	US-05-116-622-5	Sequence 5, Appli
41	71	7.1	927	4	US-09-219-277-5	Sequence 5, Appli
42	71	7.1	927	4	US-09-599-661-5	Sequence 5, Appli
43	71	7.1	2191	3	US-08-632-806A-6	Sequence 6, Appli
44	71	7.1	2192	1	US-08-273-538A-6	Sequence 6, Appli
45	70	7.0	9069	4	US-08-961-527-97	Sequence 97, Appl

ALIGNMENTS

RESULT 1

US-09-221-017B-726

; Sequence 726, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: CROSS, BRUCE C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FASTSEQ for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/221,017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1182  
; FILING DATE: 31-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP1546  
; FILING DATE: 30-JAN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP2911  
; FILING DATE: 09-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU98/01023  
; FILING DATE: 10-DEC-1998



Db 114 ATTCTAAGGCACACCGCCTACATGTCAGGTCTCTTGAAATGGATGCGCATCAGGAT 173  
QY 42 LeuProSerAspAlaAsnPhetMetArgMetSerTyrGlySerIleLysGlyTyrGlu 61  
Db 174 CAATTCCCGGATGCTAACCTGACACCTCGTTTACCTATGTCAGTGAAGGCTATTCA 233  
QY 62 ProGlnAspGlyAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuGluLysGln 81  
Db 234 CCCCGTGACAAATTTACTACGACATCAACACCATTTGGATGCTGTGTGTAAGAAAGAA 293  
QY 82 AspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLys 101  
Db 294 GATCCCGATATTGGAAATTTAGTCGATCCCAAGCTGAACCGGTATACGAGCGTAAA 353  
QY 102 AsnTyrGlyArgTyrAlaGlu---AsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsn 120  
Db 354 GACTTCGGGGTTATGCGGATCCGAGCGTGCATGCTGTAGCCTTTTCGCCACACACA 413  
QY 121 AspIleThrGlyLysSerGlySerProValPheAspLysAsnGlyArgLeuIleGly 140  
Db 414 CAVACAAACCGCGCACTCAGGAGTCCCGTTCATGAATGCCAAGCGCAACTGATCGGT 473  
QY 141 LeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeu 160  
Db 474 CTCAACTTCGATCGTAACCTGGAGGAGTCCGTCGACATCCAGTATCTGCGCGACTAC 533  
QY 161 GlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGln 180  
Db 534 CAGCGCAGCATATTGGATATTCGCTAGCTCTGCTGTATAGACAAAGTAGCGGT 593  
QY 181 CysProArgLeuIleGlnGluLysLeuLeu 191  
Db 594 TGCCAAAGCGCTGTTGGATGAATGAATATCGTT 626

## RESULT 3

US-09-058-411-1

; Sequence 1, Application US/09058411  
; Patent No. 6171854  
; GENERAL INFORMATION:  
; APPLICANT: GALLER, Ricardo  
; TITLE OF INVENTION: VACCINES AGAINST INFECTIONS CAUSED BY YF  
; TITLE OF INVENTION: VIRUS; YF INFECTIOUS CDNA, METHOD FOR PRODUCING A  
; TITLE OF INVENTION: RECOMBINANT YF VIRUS FROM THE YF INFECTIOUS CDNA AND  
; TITLE OF INVENTION: PLASMIDS TO ASSEMBLE THE YF INFECTIOUS CDNA  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBAY & CUSHMAN  
; ADDRESSEE: Intellectual Property Group of  
; ADDRESSEE: PILLSBURY MADISON & SUTRO LLP  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/058,411  
; FILING DATE: April 10, 1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: BR 9701774.4  
; FILING DATE: 11-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE JR, PAUL E  
; REGISTRATION NUMBER: 32011  
; REFERENCE/DOCKET NUMBER: 31329/251760  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944

; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10862 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: YFiv5.2/DD  
US-09-058-411-1

Alignment Scores:  
Pred. No.: 1.53 Length: 10862  
Score: 89.50 Matches: 39  
Percent Similarity: 39.22% Conservative: 21  
Best Local Similarity: 25.49% Mismatches: 46  
Query Match: 8.92% Indels: 47  
DB: 4 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-058-411-1 (1-10862)

QY 19 AlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPro 38  
Db 4742 GCITTCCTTCTGAGGAATGCCAAGAGTTG----- 4771  
QY 39 GlyArgAlaLeuProSerAspAlaAsnPhetMetArgMet---SerTyrGlySerIle 57  
Db 4772 -----ATTCCATCTTGGCTTCAGTAAAGGAACACCTTGTCGCTATGGT----- 4816  
QY 58 LysGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHis----- 71  
Db 4817 ---GGTCATCGGAAGTTGGAAGGCAGATGGCATGCGAGGAAGAGTCCAGTTGATCGCG 4873  
QY 72 ---ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaVal 90  
Db 4874 GCTGTTCCAGAAAGAACGTGGTCAACGTCCAGACAAACCGAGC----- 4918  
QY 91 GlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGly 110  
Db 4919 -----TTGTTCAAAGTGAAGAAATGGGGA-----GAAATCGGG 4951  
QY 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerPro 130  
Db 4952 GCTGTCGCTCTT-----GACTATCCGAGTGGCACTTCAGGATCTCCT 4993  
QY 131 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMet 150  
Db 4994 ATGTGTTAACAGAACGAGAGGATGTTGGCTGTACGGCAATGGC-----ATCCTT 5044  
QY 151 SerGlyAspIleGluPheGluProAspLeuGlnArgThr 163  
Db 5045 GTCGTCGACAACTCCTTCGTGTCGCGCCATATCCAGACT 5083

## RESULT 4

US-09-134-001c-1520  
; Sequence 1520, Application US/09134001c  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001c  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1520  
; LENGTH: 1083  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis

## US-09-134-001C-1520

Alignment Scores:  
Pred. No.: 0.412 Length: 1083  
Score: 83.00 Matches: 34  
Percent Similarity: 35.25% Conservative: 15  
Best Local Similarity: 24.46% Mismatches: 46  
Query Match: 8.28% Indels: 44  
DB: 4 Gaps: 6

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-134-001C-1520 (1-1083)

QY 57 IleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGly 76  
Db 58 ATGAAGGGCGCATCCCATCAAAACATCAAGAAACCCCTTTGATTAAAGGTAAGTCA 117  
QY 77 ValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 96  
Db 118 TTGCTAAAGAGTATGATCTTACAGGTGAAGAATTT-----GAAGGTCTAATCGAT 168  
QY 97 LeuPheArgThr---LysAsnTyr-----GlyArgTyrAlaGluAsn 109  
Db 169 TTTGCTATGACATTAATAAATAATAACACAGGCACACCATCGATATTAGAGGT 228  
QY 110 GlyGlnLeu----- 112  
Db 229 AAGAACATTCCTTTACTCTCGAAAGACATCTACTCGGACGCGTCCGCATTTACAGTC 288  
QY 113 -----HisIleAlaPheLeuSerAsnAsnAspIleThrGlyGly 125  
Db 289 GCATCATTTGATGATAGTCACACCCCTGAATTTTAGGGAAAAATGATATTCATAGGA 348  
QY 126 AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly 145  
Db 349 AAAAAGAACTCT-----GTTGAGGATACTGCTAAGTTTATAGGCACAACTGTTGATGA 402  
QY 146 AsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164  
Db 403 -----ATTGAATTTAGAGGTTTTTACAAAAAACTGTT 435

## RESULT 5

US-09-660-587-3  
; Sequence 3, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: 2000-09-12  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 3  
; LENGTH: 849  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; OTHER INFORMATION: nucleic acid sequence of p28-5

US-09-660-587-3  
Alignment Scores:  
Pred. No.: 0.521 Length: 849  
Score: 81.00 Matches: 43  
Percent Similarity: 39.38% Conservative: 20  
Best Local Similarity: 26.88% Mismatches: 67  
Query Match: 8.08% Indels: 30  
DB: 4 Gaps: 7

## US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-660-587-3 (1-849)

QY 23 GluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeu 42  
Db 172 GAGAAAAAAGACAACTGTAGTATATGGCTTAAAGAAAACTGGGCGAGAGATGCANTA 231  
QY 43 ProSer-----AspAlaAsnPheThrMetArg-----MetSerTyrGlySer 56  
Db 232 TCTAGTCAAAGTCCAGATGATAATTTTACCATTGCAATTTACTCATTCAAGTATGCAAGC 291  
QY 57 IleLys-----GlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGly 74  
Db 292 AACACATTTTATAGGGTTTGCAGTAGCTATTGGT-----TACTCGATAGGC 336  
QY 75 LysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIle 94  
Db 337 AGTCCAAAGATAGAA----- 363  
QY 95 LeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIle 114  
Db 364 TATGAAGCATTTTCATGTGAAAAATCCAGGTGATAATTTACAAAAACGCTGTACAGGTAT 423  
QY 115 AlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerPro----- 130  
Db 424 TGTGCTTTATCTCATCAAGATGATCGGATGATGACATGATGACACTGACAAATTT 483  
QY 131 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsn-----TrpGlu 148  
Db 484 GTATATTTAATTAATGAAGGATTACTTAACATATCATTTATGACAAACATATGTTATGAA 543  
QY 149 AlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIle 168  
Db 544 ACAGCAAGCAAAATATATACCTCTCTCTCATATATGTCAGGTATTGCTACTGATTTA 603

## RESULT 6

US-09-261-358A-3  
; Sequence 3, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261,358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201,458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 3  
; LENGTH: 849  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; OTHER INFORMATION: nucleic acid sequence of Eca28sA2

US-09-261-358A-3  
Alignment Scores:  
Pred. No.: 0.521 Length: 849  
Score: 81.00 Matches: 43  
Percent Similarity: 39.38% Conservative: 20  
Best Local Similarity: 26.88% Mismatches: 67  
Query Match: 8.08% Indels: 30  
DB: 4 Gaps: 7

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-261-358A-3 (1-849)

QY 23 GluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeu 42  
Db 172 GAGAAAAAAGACAACTGTAGTATATGGCTTAAAGAAAACTGGGCGAGAGATGCANTA 231

```

Db      378 AGCAACTACAAAATCCCGCTGCATGCTCGGGCAAGCTGCATTTCGGTGTC A----- 428
Qy      78 LeuGlutylsGlnAspProLysSerAspGluPheAlaValcIcnGluAsnIleLeuAspLeu 97
Db      429 -----GCCGTGGCGAGCAGTACATCGACCTG 455
Qy      98 PheArgThrLysAsnTyrglyArgtyrAlaGluAsnGlycInLeuHisIleAlaPheLeu 117
Db      456 GTGTCCACC GGCTGCTCGGGTAATACTTCTCTCC TCCGAG C----- 497
Qy      118 SerAsnAsnAspIlethrGlyGlyAsnSerGlySer-----ProValPheAspLys 134
Db      498 -----ACCATCACACAGGACCGTTTCCCAGTCAGATCGGCGCGCTGGACAAT 548
Qy      135 AsnGlyArg-----LeuIleGlyLeuAlaPheAspGlyAsn 146
Db      549 TCCATCTCGGGTGGCGCATTCCTCCACGAGAAGATCGCTGTGCTCGTAGCAGAC 608
Qy      147 TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164
Db      609 GCGCAACCGGTGGGTGG-----CTGGGACCCGCGTTCGAACGGTTGGTC 653

RESULT 8
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:          1.63e+05           Length:       4403765
Score:             77.50              Matches:        40
Percent Similarity: 32.91%            Conservative:   12
Best Local Similarity: 25.32%         Mismatches:    45
Query Match:       7.73%              Indels:        61
DB:                4                  Gaps:          8

US-10-008-355-2_COPY_522_712 (1-191) x US-09-103-840A-2 (1-4403765)
Qy      36 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArg---MetSerTyr 54
Db      205549 CTGTATCCG-----ACGGCCAATGTGACCTACCGCGGTATCACCAATT 205590
Qy      55 GlySerIleLysGlyTyrglyProGlnAsp----- 64
Db      205591 GGCAAGGTACTCGCTCGCAGCCCCACCGCAGGGCGCAGAGTGACGATGACATCGCC 205650
Qy      65 -----GlyAlaTrpTyraAsnTyrrHisThrThrGlyLysGlyVal 77
Db      205651 AGCAACTACAAAATCCCGCTGCATGCGCTCGCGNACGTGCAATTCGGTGTC A----- 205701
Qy      78 LeuGlutylsGlnAspProLysSerAspGluPheAlaValcIcnGluAsnIleLeuAspLeu 97

```

Db	205702	-----CCGGTGGGCGAGCAGTACATCGACCTG	205720
QY	98	PheArgThrLysAsnTy rGlyArGTyrAlaGluAsnGlycInLeuHisIleAlaPheLeu	117
Db	205729	GTGTCACCGGTCGCCGGTAATAACTTCTCCTCCGGACAG-:::	205770
QY	118	SerAsnAsnAspIlethrGlyGlyAsnSerGlySer-----ProValPheAspLys	134
Db	205771	-----ACCATCACCAAGGCACCGTTGCCAGTGAGATCGGCGCGCTGGACAAT	205821
QY	135	AsnGlyArg-----LeuIleGlyLeuAlaPheAspGlyAsn	146
Db	205822	TCCAATCGCGGTGGCCGCATTGCCACGAGAGAAGATCGGTGCTGTGCAGCAGAGACC	205881
QY	147	TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle	164
Db	205882	GCCCAACCGTGGGTGGG-----CTGGGACCCGCGTTGCAACGGTGTGTC	205926
RESULT 9			
US-09-103-840A-1			
; Sequence 1, Application US/09103840A			
; Patent No. 6294328			
; GENERAL INFORMATION:			
; APPLICANT: FLEISCHMAN, Robert D.			
; APPLICANT: WHITE, Owen R.			
; APPLICANT: FRASER, Claire M.			
; APPLICANT: VENTER, John C.			
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM			
; FILE REFERENCE: TUBERCULOSIS			
; CURRENT APPLICATION NUMBER: US/09/103,840A			
; CURRENT FILING DATE: 1998-06-24			
; NUMBER OF SEQ ID NOS: 2			
; SOFTWARE: Patentin Ver. 2.1			
; SEQ ID NO 1			
; LENGTH: 4411529			
; TYPE: DNA			
; ORGANISM: Mycobacterium tuberculosis			
; OTHER INFORMATION: H37Rv			
US-09-103-840A-1			
Alignment Scores:			
Pred. No.: 1.63e+05 Length: 4411529			
Score: 77.50 Matches: 40			
Percent Similarity: 32.91% Conservative: 12			
Best Local Similarity: 25.32% Mismatches: 45			
Query Match: 7.73% Indels: 61			
DB: 4 Gaps: 8			
US-10-008-355-2_COPY_522_712 (1-191) x US-09-103-840A-1 (1-4411529)			
QY	36	MetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArg---MetSerTyr	54
Db	205382	CTGTATCCG-----ACGGCCAATGTGACCTACCGCGGTATCCACCAT	205423
QY	55	GlySerIleLysGlyTyrGluProGlnAsp-----	64
Db	205424	GGCAAGGTACTCCGTCGAGCCCCACGACCGAGGGCCAGCAGTACGATGACCATCGCC	205483
QY	65	-----GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyVal	77
Db	205484	AGCAACTACAAATCCCGCTCGATGCTCGCGCAACGTGTCATTCGGTGTCA-----	205534
QY	78	LeuGlutLysGlnAspproLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu	97
Db	205535	-----GCGGTGGCGAGCAGTACATCGACCTG	205561
QY	98	PheArgThrLysAsnTy rGlyArGTyrAlaGluAsnGlycInLeuHisIleAlaPheLeu	117
Db	205562	GTGTCACCGGTCGCCGGTAATAACTTCTCCTCCGGACAG-----	205603
QY	118	SerAsnAsnAspIlethrGlyGlyAsnSerGlySer-----ProValPheAspLys	134

```

Db 205604 -----ACCATCAACGAAGGCACCGTTTCCACAGTTGAGATCGCGCGCGCTGGACAAT 2056655
Qy 135 AsnGlyArg-----LeuIleGlyLeuAlaPheAspGlyAsn 146
   ::: :||| :|||
Db 205655 TCCAATCGCGGTTGGCCGCATTCGCCACGAGAGAAGATCGCTTGCTGCTGCAGCAGAC 2057114
   :|||
Qy 147 TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164
   :::|||:: :|||
Db 205715 GCGCAAGCGGTGGTGGG-----CTGGGACCGCGCTGCAACGGTTGGTTC 205759

RESULT 10
US-09-441-340-2
; Sequence 2, Application US/09441340.
; Patent No. 648476
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; TITLE OF INVENTION: Phosphonate Metabolizing Plants
; FILE REFERENCE: 38-21(15303)
; CURRENT APPLICATION NUMBER: US/09/441,340
; CURRENT FILING DATE: 1999-11-16
; EARLIER APPLICATION NUMBER: 60/108,763
; EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 11672
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-441-340-2

Alignment Scores:
Pred. No.: 106
Score: 75.00
Percent Similarity: 40.43%
Best Local Similarity: 29.79%
Query Match: 7.48%
DB: 4

US-10-008-355-2_COPY_522_712 (1-191) x US-09-441-340-2 (1-11672)
Qy 5 IleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGlu--- 23
   ||| ||| :|||:::
Db 1034 ATCAGCAGCAAACTGAAACCGCAATGAGCGCGCTTCTGCAGGATATGGAGAAAGACT 1093
Qy 24 -----LysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyA-ArgAla 41
   :::|||:::||||| :|||
Db 1094 GGGCGTGAAGGTCAACGCCCTCTTTGGCCCGGACTACGCCGGGCAATTATCCAGGGGATGCG 1153
Qy 42 LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu 61
   ||| :||| :|||
Db 1154 CTTCAATAAAGTGGATATCGCTGTGACGCAATCTCTCGCGCATGGAACGGTGGATCG 1213
Qy 62 ProGlnAspGlyAlaTyrTyrAsnTyrHisThrThrGlyLysGly----- 76
   ||| ||| :|||
Db 1214 CGCCAA-----TGCCCAAGGTCTTTCGCCCAAGACAGCGTCGCGCGGATGATGATCGCGG 1264
Qy 77 -----ValLeuGluLysGlnAspProLysSerAspGlu 87
Db 1265 TTAGTGGAGCGTGTGTATGCTGTCACAAAGACAGTCCGATCAA 1306

RESULT 11
US-09-454-721A-3
; Sequence 3, Application US/09454721A
; Patent No. 6296854
; GENERAL INFORMATION:
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Burce J. Crise
; TITLE OF INVENTION: Live Attenuated Venezuelan Equine Encephalitis Vaccin
; FILE REFERENCE: Army 146
; CURRENT APPLICATION NUMBER: US/09/454, 721A
; CURRENT FILING DATE: 1999-12-07

```







; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2989 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 390..2768  
; US-08-942-046-9

Alignment Scores:  
Pred. No.: 21.6 Length: 2989  
Score: 74.00 Matches: 39  
Percent Similarity: 33.88% Conservative: 23  
Best Local Similarity: 21.31% Mismatches: 55  
Query Match: 7.38% Indels: 66  
DB: 4 Gaps: 9

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-08-942-046-9 (1-2989)

QY 16 MetAlaAsnAlaTyrAlaIleGluLysGlyArgLeuPhePheAlaGlyLeuArgGlu 35  
Db 1383 TTACGGATACACCTTTGTTGTATGCTGGACGAGCTTAACCTTCGCCAACCTCGC--- 1439  
QY 36 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 55  
Db 1440 ---TTTGAAGGAATACCGTTCTGCTGTAGTAGTATTACGTCAGGAATG----- 1487  
QY 56 SerIleLysGlyTyrGluProGlnAspGlyAlaTyrPyrAsnTyrHisThrThrGlyLys 75  
Db 1488 -----CGACACCAAGAGGAACTTGGTATATTAATTCACAATTAGTTGAGTTA 1532  
QY 76 GlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeu 95  
Db 1533 GGAATAATTCGCTTAGAT-----CGTACAGGTTTCTTCGAACAGTTGAAACCGAATT 1586  
QY 96 Asp-----LeuPheArgThrLys----- 101  
Db 1587 GATCCTATCAATGGTAGCAATGATGAAGTGGATGCTCGTATATAAAGTCAAGAACGTAAC 1646  
QY 102 -----AsnTyrGlyArgTyrAlaGluAsnGly----- 110  
Db 1647 ACGGGTAGTATCAACTTTGGTATTGGTTACGGT-----ACAGAGAGTGGTATCAGTTAT 1700  
QY 111 -----GlnLeuHisIleAlaPhe 116  
Db 1701 CAACAAGTATTAAACAAGATAATTTCTTGGACAGCGCGCAGTAAGTATAGCTGGT 1760  
QY 117 LeuSerAsnAsnAspIleThrGlyLysAsnSerGly-----SerProValPheAspLys 134  
Db 1761 ACGAAAATGATTATGTCAGGAGTGTCAATTTGGTTATACCGAACCCCTATTTACTAAA 1820  
QY 135 AsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIle 154  
Db 1821 GATGGGTGAAGTCTTGGT-----GGAAATATT 1847  
QY 155 GluPheGlu 157  
Db 1848 TTCTTTGAA 1856

RESULT 15

US-09-206-942-52  
; Sequence 52, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,568  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 52  
; LENGTH: 2934  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; US-09-206-942-52

Alignment Scores:  
Pred. No.: 24.3 Length: 2934  
Score: 73.50 Matches: 40  
Percent Similarity: 37.14% Conservative: 12  
Best Local Similarity: 28.57% Mismatches: 69  
Query Match: 7.33% Indels: 19  
DB: 4 Gaps: 6

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-206-942-52 (1-2934)

QY 42 LeuProSerAspAlaAsnPheThrMetArgMet-----SerTyrGlySerIleLys 58  
Db 631 TTAAAGCAATCATCGGAATTTACCTTTATTAATATATCGACAGCAATCATTCACAAAC 690  
QY 59 GlyTyrGluProGlnAspGlyAlaTyrPyrAsnTyrHisThrThrGlyLysGlyValLeu 78  
Db 691 TCCAGTGCATTACACGAAGCTTTTCGGGGAGTAAAGTTCCAC-----GCCAAGAAAT 738  
QY 79 GluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPhe 98  
Db 739 -----AATGAAATGAAATTTAATATTGTTAATAATGCCAAGGCTGAATTT 783  
QY 99 ArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSer 118  
Db 784 AGTTAAACCAATGAGAGACAACTCCCTAACAGACCACTACCACAAATCAGTTTTTATCT 843  
QY 119 AsnAsnAspIleThrGlyLysAsnSerGlySerProValPheAspLysAsnGlyArgLeu 138  
Db 844 AATATTTCCGTCCTACTGCGCGA-----GGTCTGTGTTTTTCGATATATACGCTAACCTT 897  
QY 139 IleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluPro 158  
Db 898 TGGGGT-----AAAGGAGCTGAGCTAAAGATGGATTCAATTAACGTTTCTAGC 945  
QY 159 AspLeuGlnArgThrIleSerValAspIleArg---TyrValLeuPheMetIleAspLys 177  
Db 946 GGCTCTAATCTTACCTTAATTCCTCCATGTTCCGCAAGTATATGCTTTTGAATCAATAA 1005

Search completed: May 23, 2003, 14:01:23  
Job time : 1702.29 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 10:12:58 ; Search time 50.1296 Seconds  
(without alignments)

5031.128 Million cell updates/sec

Title: us-10-008-355-2\_COPY\_522\_712

Perfect score: 1003

Sequence: 1 SXSVTAAARAIQADAMANY.....LFMDKWGCQPRLOELKLI 191

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110406\_2649/app:query.fasta\_1.1230  
-DB=Published Applications\_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10008355 -ECGN\_1\_1\_132 -runat\_16052003\_110406\_2649  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1003	100.0	2139	9	US-10-008-355-1
2	85.5	8.5	651	10	US-09-770-149-525
3	81.5	8.1	720	10	US-09-815-242-4296
4	81.5	8.1	765	10	US-09-815-242-8179
					Sequence 1, Appli
					Sequence 525, App
					Sequence 4296, Ap
					Sequence 8179, Ap

5	81.5	8.1	1812	9	US-09-938-842A-2020	Sequence 2020, Ap
6	81	8.1	849	9	US-10-062-051-3	Sequence 3, Appli
7	81	8.1	849	12	US-10-062-624-3	Sequence 3, Appli
8	80	8.0	1019	7	US-08-781-986A-752	Sequence 752, App
9	79.5	7.9	861	10	US-09-974-300-2760	Sequence 2760, Ap
10	78.5	7.8	6444	7	US-08-781-986A-233	Sequence 233, App
11	78	7.8	525	9	US-09-981-286A-6	Sequence 6, Appli
12	77.5	7.7	503	9	US-09-981-286A-31	Sequence 31, Appli
13	77.5	7.7	841	9	US-09-996-634-39	Sequence 39, Appli
14	77.5	7.7	841	9	US-09-997-181-39	Sequence 39, Appli
15	77.5	7.7	841	9	US-09-997-182-39	Sequence 39, Appli
16	77	7.7	1238	7	US-08-781-986A-232	Sequence 232, App
17	76.5	7.6	1026	10	US-09-815-242-7701	Sequence 7701, Ap
18	76	7.6	888	10	US-09-974-300-1729	Sequence 1729, Ap
19	76	7.6	3163	9	US-09-822-846-213	Sequence 213, App
20	76	7.6	4474	9	US-09-909-567B-7	Sequence 7, Appli
21	75	7.5	1986	10	US-09-974-300-1841	Sequence 1841, Ap
22	75	7.5	3280	9	US-10-101-464A-884	Sequence 884, App
23	74.5	7.4	1026	10	US-09-991-258-9	Sequence 9, Appli
24	74.5	7.4	5076	10	US-09-991-258-8	Sequence 8, Appli
25	74.5	7.4	14759	10	US-09-952-360-1	Sequence 1, Appli
26	74	7.4	478	9	US-09-918-995-16065	Sequence 16065, A
27	74	7.4	1033	9	US-10-101-464A-413	Sequence 413, App
28	74	7.4	2604	9	US-10-101-464A-834	Sequence 834, App
29	73.5	7.3	477	9	US-09-981-286A-12	Sequence 12, Appli
30	73.5	7.3	1472	10	US-09-974-300-1065	Sequence 1065, Ap
31	73	7.3	834	9	US-09-738-626-1216	Sequence 1216, Ap
32	73	7.3	1282	9	US-09-987-107-43	Sequence 43, Appli
33	73	7.3	1311	10	US-09-815-242-7057	Sequence 7057, Ap
34	73	7.3	3170	9	US-10-135-322-4	Sequence 4, Appli
35	73	7.3	3170	9	US-10-135-322-23	Sequence 23, Appli
36	73	7.3	3174	9	US-09-918-508-5	Sequence 5, Appli
37	73	7.3	3503	9	US-10-135-322-2	Sequence 2, Appli
38	73	7.3	3612	9	US-10-135-322-3	Sequence 3, Appli
39	73	7.3	3620	9	US-10-135-322-1	Sequence 1, Appli
40	73	7.3	11517	10	US-09-901-106-1	Sequence 1, Appli
41	73	7.3	70768	9	US-10-135-322-13	Sequence 13, Appli
42	72.5	7.2	842	12	US-10-062-994-3	Sequence 3, Appli
43	72.5	7.2	842	12	US-10-062-994-3	Sequence 3, Appli
44	72.5	7.2	1188	9	US-09-938-842A-615	Sequence 615, App
45	72.5	7.2	1285	9	US-09-987-107-45	Sequence 45, Appli

ALIGNMENTS

RESULT 1  
US-10-008-355-1  
; Sequence 1, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 2139  
; TYPE: DNA  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-1

Alignment Scores:			
Pred. No.:	2,91e-127	Length:	2139
Score:	1003.00	Matches:	191
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0



; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4296  
; LENGTH: 720  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-4296

Alignment Scores:  
Pred. No.: 0.111 Length: 720  
Score: 81.50 Matches: 34  
Percent Similarity: 37.84% Conservative: 22  
Best Local Similarity: 22.97% Mismatches: 59  
Query Match: 8.13% Indels: 33  
DB: 10 Gaps: 6

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-815-242-4296 (1-720)

QY 37 TyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySer 56  
Db 322 TATCCAGGTAAGAGATGATCATCTCAAGTTGAAGAGCGTGCAATAGACGGTGGT 381  
QY 57 IleLysGlyTyrGluProGlnAspGlyAlaTrp---TyrAsnTyrHisThrThrGlyLys 75  
Db 382 CCAAAAGGATTTAATTTAATGATAATGTAACGCCGTTCAAAATATGCAGCAGGGGCTAAA 441  
QY 76 Gly-----ValLeuGluLysGlnAspProLysSerAspGluPheAlaVal 90  
Db 442 GCTGGTGACCGAATTAAGTATGATCGGTATCCACACCCATACAAAATAATATGTTT 501  
QY 91 GlnGluAsnIleLeuAspLeuPheArgThrLysAsn-----TyrGlyArgTyr 106  
Db 502 CATGAGTCAACTGCTGCTGATCTCAGTAGAAGGTAGCAGCATGTTATATTTCAGCCCAT 561  
QY 107 AlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsn 126  
Db 562 ACTGAAAGT-----GAAAC 576  
QY 127 SerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe----- 143  
Db 577 TCTGGATCAGCTGTTAATTAACAGCAACCAAGTAATATAGTATTTATTTGCTTCTCAT 636  
QY 144 -----AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeu 160  
Db 637 GTAAAAATGATGACACAGAAATGATATGCG-----GTCTACTTTACACCAAGAAATT 690  
QY 161 GlnArgThrIleSerValAspIle 168  
Db 691 AAAAAGTTCATTGCGAAGAAACATA 714

## RESULT 4

US-09-815-242-8179  
; Sequence 8179, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlson, Kari L.  
; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8179  
; LENGTH: 765  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(765)  
US-09-815-242-8179

Alignment Scores:  
Pred. No.: 0.121 Length: 765  
Score: 81.50 Matches: 34  
Percent Similarity: 37.84% Conservative: 22  
Best Local Similarity: 22.97% Mismatches: 59  
Query Match: 8.13% Indels: 33  
DB: 10 Gaps: 6

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-815-242-8179 (1-765)

QY 37 TyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySer 56  
Db 364 TATCCAGGTAAGAGATGATCATCTCAAGTTGAAGAGCGTGCAATAGACGGTGGT 423  
QY 57 IleLysGlyTyrGluProGlnAspGlyAlaTrp---TyrAsnTyrHisThrThrGlyLys 75  
Db 424 CCAAAAGGATTTAATTTAATGATAATGTAACGCCGTTCAAAATATGCAGCAGGGGCTAAA 483  
QY 76 Gly-----ValLeuGluLysGlnAspProLysSerAspGluPheAlaVal 90  
Db 484 GCTGGTGACCGAATTAAGTATGATCGGTATCCACACCCATACAAAATAATATGTTT 543  
QY 91 GlnGluAsnIleLeuAspLeuPheArgThrLysAsn-----TyrGlyArgTyr 106  
Db 544 CATGAGTCAACTGCTGCTGATCTCAGTAGAAGGTAGCAGCATGTTATATTTCAGCCCAT 603  
QY 107 AlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsn 126  
Db 604 ACTGAAAGT-----GAAAC 618  
QY 127 SerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe----- 143  
Db 619 TCTGGATCAGCTGTTAATTAACAGCAACCAAGTAATATAGTATTTATTTGCTTCTCAT 678  
QY 144 -----AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeu 160  
Db 679 GTAAAAATGATGACACAGAAATGATATGCG-----GTCTACTTTACACCAAGAAATT 732

Qy 161 GlnArgThrIleSerValAspIle 168  
Db 733 AAAAAGTTTCATTCGAGAAACATA 756

## RESULT 5

US-09-938-842A-2020  
; Sequence 2020, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIPT1300-3  
; CURRENT APPLICATION NUMBER: US/09/938, 842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 2020  
; LENGTH: 1812  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-2020

Alignment Scores:  
Pred. No.: 0.416 Length: 1812  
Score: 81.50 Matches: 53  
Percent Similarity: 38.05% Conservative: 33  
Best Local Similarity: 23.45% Mismatches: 67  
Query Match: 8.13% Indels: 74  
DB: 13 Gaps: 13

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-938-842A-2020 (1-1812)

Qy 18 AsnAlaTyr-----AlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArg--- 34  
Db 421 AATCTCTACGTCCGGGATAGACAAAGGAGGTTATCTTCTTATTCGTGAAGTCG 480  
Qy 35 GluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThr-----MetArgMet 52  
Db 481 GAGTCAAGACGTCAGGTGGGTACCGGCTCGCCGGTTCTCAGGAGCTTATTACAAAGC 540  
Qy 53 SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr 72  
Db 541 GACCACCTCAAGAGACGACGTCAGATCCG-----TACAACGCTACACT 585  
Qy 73 ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAla----- 89  
Db 586 AGTCTTAACGAAGCCACTCTCTGCTCCGACTCGTCCCAAGCATGTATGCTCAAAATGCTA 645  
Qy 90 -----ValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArg 105  
Db 646 TGTGGTCTCTAATGCCCATGAAGTTCTCCGACTC-----GGCGCA 687  
Qy 106 TyrAlaGluAsnGlyGlnLeuHis---IleAlaPheLeuSerAsnAsn----- 120  
Db 688 GTGTTTCTCTCCGCTCCCTCCGTCGCATTAAGCTTCCTCCAGAACAAATTGGAAGAACTT 747  
Qy 121 -----AspIleThrGlyLysAsnSerGlySerProValPheAsp-----LysAsn 135  
Db 748 GCTCGTGATATCTCAACCGGAACCCCTAAGTTCTCGAATCTCTCGATCTCGCATTAATAAC 807  
Qy 136 GlyArg-----LeuIleGlyLeu 141  
Db 808 CGCATGTCCAAGATTTTGACCAACACCTGATCAAGAACTGGCTGAGTTTTTGGTTGGGGTT 867

Qy 142 AlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIle----- 154  
Db 868 TGTTCACAAGAAATTGGGAAGGATATACAAAGATATGGCTAAACACGAGTACCTC 927  
Qy 155 -----GluPheGluProAspLeuGlnArgThrIleSer 165  
Db 928 GACGTGATGTTACTGGAGCAATGGCTCAGTATATCCGAGCTTGGAGT---ACTATAGCG 985  
Qy 166 ValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCys-----Pro 182  
Db 986 GTGATTACCGATGGCTT-----GCACGATGATGCTTCCTCGTCCG 1024  
Qy 183 ArgLeuIleGlnGluLeu 188  
Db 1025 AAAGTTATTTCGGGATTA 1042  
RESULT 6  
US-10-062-051-3  
; Sequence 3, Application US/10062051  
; Publication No. US20030073095A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/10/062,051  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US/09/660,587  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261,358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 3  
; LENGTH: 849  
; TYPE: DNA  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; OTHER INFORMATION: nucleic acid sequence of p28-5  
US-10-062-051-3  
Alignment Scores:  
Pred. No.: 0.165 Length: 849  
Score: 81.00 Matches: 43  
Percent Similarity: 39.38% Conservative: 20  
Best Local Similarity: 26.88% Mismatches: 67  
Query Match: 8.08% Indels: 30  
DB: 9 Gaps: 7  
US-10-008-355-2\_COPY\_522\_712 (1-191) x US-10-062-051-3 (1-849)  
Qy 23 GluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeu 42  
Db 172 GAGAAAAAAGACAACTAGTAGTATATGGCTTAAAGAAAACTGGCGAGAGATGCAATA 231  
Qy 43 ProSer-----AspAlaAsnPheThrMetArg-----MetSerTyrGlySer 56  
Db 232 TCTAGTCAAAAGTCCAGATGATAATTTCACCATTCGAGTAGCTATTGGT-----TACTCGATAGGC 291  
Qy 57 IleLys-----GlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGly 74  
Db 292 AACAAAGTTTTCAGGTTTTCAGTAGCTATTGGT-----TACTCGATAGGC 336  
Qy 75 LysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIle 94  
Db 337 AGTCCAAAGAATAAGAA-----GTGAGATGCTCT 363  
Qy 95 LeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIle 114  
Db 364 TATGAACGATTTGATGTGAAAAATCCAGGTGATTAATTACAAAAACGGTGTCTACAGGTAT 423



```

Qy 115 AlapheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerPro----- 130
Db 424 TGTGCTTTATCTCATCAAGATGATGCGGATGATGACATGCTCAACTGACAAATTT 483
Qy 131 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlapheAspGlyAsn-----TrpGlu 148
Db 484 GTATATTAAATTAATGAAGGATTACTTAACATATCATTTATGACAAACATATGTTATGAA 543
Qy 149 AlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIle 168
Db 544 ACAGCAAGCAAAAATATACCTCTCTCTCTTACATATGTGCAGGTATGTGCTACTGATTGA 603

RESULT 7
US-10-062-624-3
; Sequence 3, Application US/10062624
; Patent No. US20020115840A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2/D1
; CURRENT APPLICATION NUMBER: US/10/062,624
; PRIORITY FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/660,587
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 3
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of p28-5
US-10-062-624-3

```

[illegible]

```

Db 484 GTATATTTAATTAATGAAGATTACTTACATATCATCTTTATGACAAACATATGTTATGAA 543
Qy 149 AlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIle 168
Db 544 ACAGCAAGCAAAAATATACCTCTCTCTCTACATATGTGCAGGTATTGCTACTGATTTA 603

RESULT 8
US-08-781-986A-752
; Sequence 752, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 752:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1019 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-752

Alignment Scores:
Pred. No.: 0.293 Length: 1019
Score: 80.00 Matches: 17
Percent Similarity: 55.56% Conservative: 13
Best Local Similarity: 31.48% Mismatches: 24
Query Match: 7.98% Indels: 0
DB: 7 Gaps: 0

US-10-008-355-2_COPY_522_712 (1-191) x US-08-781-986A-752 (1-1019)
Qy 123 ThrGlycIAsnSerGlySerProValPheAspIleGlyLeuAla 142
Db 642 ACTGGTGGTATTTTTCACCTGTATTAATGAAAAAAGAGTGCAGGATTCAT 701
Qy 143 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 162
Db 702 TGGGGCGGTGTACCAATGAATTAATGGGCGCGTATTATTAAATGAAATGTACGCAAC 761
Qy 163 ThrIleSerValAspIleArgTyrValLeuPheMetIleAsp 176
Db 762 TTCITAAACAAAATATTTGAAGATATCCATTTTGCCCAACGAT 803

RESULT 9
US-09-974-300-2760
; Sequence 2760, Application US/09974300

```

; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2760  
; LENGTH: 861  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-2760

Alignment Scores:  
Pred. No.: 0.27 Length: 861  
Score: 79.50 Matches: 37  
Percent Similarity: 36.84% Conservative: 19  
Best Local Similarity: 24.34% Mismatches: 39  
Query Match: 7.93% Indels: 57  
DB: 10 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-974-300-2760 (1-861)

Qy 43 ProSerAspAlaAsnPhetMetArgMetSerTyrGlySerIleLysGlyTyrGluPro 62  
Db 455 CCITCGGATGGCTTTATCTATACGGTTAGATTATGGAACA-----AAATCCGTT 505  
Qy 63 GlnAspGlyAlaTyrAsnTyrHisThrGlyLysGlyValLeuGluLysGlnAsp 82  
Db 506 ACAACCGCTGTCGAATATGTTATGATAAAGGTTGAAGGCTATGTGAGAAAGTCAGAC 565  
Qy 83 Pro-----LysSerAspGluPheAlaValGlnGluAsnIleLeu--- 95  
Db 566 GCGGCTTTTGACACAGACCTTGAAACAGAGAACCGGTTGTTCTTCAAAACATCTTAATA 625  
Qy 96 -----AspLeuPheArg 99  
Db 626 GCGGAGCCGCAACATCATATGAAAGATGCGTACGGGAGAGGGATATTGACCTT---CGT 682  
Qy 100 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyClnLeuHisIleAlaPheLeuSerAsn 119  
Db 683 TCAGCGGAAAGGACCTCTCTGCAAAACGGGAGAGTCCATCGGATCAGCTGGAAGAAC 742  
Qy 120 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 139  
Db 743 -----GAAACGCGCGGATCATA 760  
Qy 140 GlyLeuAlaPheAspGly-----AsnTrpGluAlaMet 150  
Db 761 CCGACGCGC---GACGGCAAAACGGTTCCGTTCTGTCGGGAAACCGTGG----- 808  
Qy 151 SerGlyAspIleGluPheGluProAspLeuGluArg 162  
Db 809 -----ATCAATATCGTTCCCGATCTCAGCAAG 835

RESULT 10  
US-08-781-986A-233/c  
; Sequence 233, Application US/08781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 233:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6444 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-781-986A-233

Alignment Scores:  
Pred. No.: 6.57 Length: 6444  
Score: 78.50 Matches: 49  
Percent Similarity: 34.08% Conservative: 27  
Best Local Similarity: 21.97% Mismatches: 66  
Query Match: 7.83% Indels: 81  
DB: 7 Gaps: 10

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-08-781-986A-233 (1-6444)

Qy 6 AlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGly 25  
Db 681 GCAGCACATGAAATTTCTCAACCTTAATATGATCGGATTAATACGCTCAAGAACAAAGG 622  
Qy 26 LysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAsp 45  
Db 621 ATAACGGTTGTTATCGGACAGGTCGAGCATTTTATGAAGCMCAAGCACCAGTTGCTGAC 562  
Qy 46 AlaAsnPhetMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp--- 64  
Db 561 ACAGATTAAACA-----GTACCATATATTGTTGAATGGTGTGCTGAAGTACGATGAA 508  
Qy 64 ----- 64  
Db 507 ACTTTCAATGTAATGAGCACTTCACACCTTAATAAATCGTTAGTACACAAAATTTACAAT 448  
Qy 65 -----GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 79  
Db 447 GTTTTAAAGATGAGGATATTATTATCAAGTATACAGAGT-----CGTGGCATTAT 394  
Qy 80 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPhe--- 98  
Db 393 ACTGAACATCCACAAAGAGAT-----TTAGACATTATACATA 358  
Qy 99 -----ArgThrLysAsn---TyrGly 104  
Db 357 GATATTGCTGAGCGTGCAGGTCACATGCAACACGTTGAGCGGTATTAATAATGGTATTCAA 298  
Qy 105 ArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly 124  
Db 297 AGACGCATAGATATGTTACGTGTCGATGAGATT-----GTTGATATATGATGCTATTGAA 244

Db	334	AAAGCAGTTGGGGCCAAAGGAGACAGCGGAGACCCCATCTCGATACCAAGGACCGGGTG	339
Qy	139	IleGlyLeuAlaPheAspGly----	
Db	394	GTCGCTATTGTGTGGAGGTGTGAATCAAGAGTCTAGGACAGCCCTTTTCAGTCGTCATG	453
Qy	147	TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerVal	166
Db	454	TGGACAAAGCTTGGATCTTCTCTCGAG-----GGAGTT	486
Qy	167	AspIleArgTyrValLeuPheMetIleAspLysTrp	178
Db	487	ACCGTGAAGTATACTCGGAGAACTGCGAGCAATGG	522
RESULT 12			
US-09-981-286A-31			
; Sequence 31, Application US/09981286A			
; Publication No. US20020192799A1			
; GENERAL INFORMATION:			
; APPLICANT: Watowich, Stanley J.			
; APPLICANT: Weaver, Scott C.			
; APPLICANT: Davey, Robert A.			
; TITLE OF INVENTION: Drug Discovery Methods			
; FILE REFERENCE: 265.00260101			
; CURRENT APPLICATION NUMBER: US/09/981,286A			
; CURRENT FILING DATE: 2001-10-15			
; PRIOR APPLICATION NUMBER: US 60/240,187			
; PRIOR FILING DATE: 2000-10-13			
; NUMBER OF SEQ ID NOS: 36			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 31			
; LENGTH: 503			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: AdapteIn-1 nucleotide sequence			
US-09-981-286A-31			
Alignment Scores:			
Pred. No.: 0.235 Length: 503			
Score: 77.50 Matches: 42			
Percent Similarity: 37.21% Conservative: 22			
Best Local Similarity: 24.42% Mismatches: 63			
Query Match: 7.73% Indels: 45			
DB: 9 Gaps: 11			
US-10-008-355-2_COPY_522_712 (1-191) x US-09-981-286A-31 (1-503)			
Qy	42	LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu	61
Db	10	TTGGAATCTGCAAGACGTTCCCAATC---ATGTTGGAAGGAAGATAAACGGCTACGCT	66
Qy	62	ProGlnAspGlyAla-----TrpTyrAsnTyrHisThrGlyLys-----	75
Db	67	TGTGTGTCGGAGGGAAGTTATTATTCAGGCCGATGTCATGTGAAGGCAAGATCGACACGAC	126
Qy	76	-----GlyValLeuGluLysGlnAspProLysSerAsp---GluPheAla---Val	90
Db	127	GTTCTGGCCGCGCTTAGACGAAAGAAAGATCCAAATACCATCTTGAGTATGCAGATGTG	186
Qy	91	GlnGluAsnIle---LeuAspLeuPheArg-----ThrLysAsnTyrGlyArgTyr	106
Db	187	CCACAGAACATCGCGCGCATACATTCAAATACACCCATGAGAAACCCCAAGGCTATTAC	246
Qy	107	Ala-----GluAsnGlyGlnLeuHisIleAlaPheLeuSer	118
Db	247	AGCTGGCATCATGAGCAGTCCAATATGAAATGGCGTTTCACGGTG-----CCG	297
Qy	119	AsnAsnAspIleThrGlyGlyAsnSerClySerProValPheAspLysAsnGlyArgLeu	138
Db	298	AAAGCAGTTGGGGCCAAAGGAGACAGCGGACGACCCCAATCTGGATAACCAAGGACGGGTG	357
Qy	139	IleGlyLeuAlaPheAspGly-----	146

Db 358 GTCCGATTGCTGGGAGGTGTAATGAAGGATCTAGGACAGCCCTTCAGTCGTCATG 417  
Qy 147 TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerVal 166  
Db 418 TGGAAACAGCTTTCT-----CCACATTATGCTCAACTCGAGGAGTT 459  
Qy 167 AspIleArgTyrValLeuPheMetIleAspLysTrp 178  
Db 460 ACCGTGAAGTATATCTCCGGAGAACTCGGAGCAATGG 495

## RESULT 13

US-09-996-634-39  
; Sequence 39, Application US/09996634  
; Patent No. US20020172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: immunostimulatory Peptides  
; FILE REFERENCE: 61260  
; CURRENT APPLICATION NUMBER: US/09/996,634  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/447,135  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08/990,823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 39  
; LENGTH: 841  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(841)  
; OTHER INFORMATION: n is a, c, g, or t/u.  
US-09-996-634-39

Alignment Scores:  
Pred. No.: 0.491 Length: 841  
Score: 77.50 Matches: 40  
Percent Similarity: 32.91% Conservative: 12  
Best Local Similarity: 25.32% Mismatches: 45  
Query Match: 7.73% Indels: 61  
DB: 9 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-996-634-39 (1-841)

Qy 36 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArg---MetSerTyr 54  
Db 276 CTGATTCG-----ACGCCAATGTGACCTACCGGGGTATCACCATT 317  
Qy 55 GlySerIleLysGlyTyrGluProGlnAsp----- 64  
Db 318 GGCAAGGTTACTGCGTCGAGCCACCACCGAGGCGCCAGTGACGATGAGCATGCC 377  
Qy 65 -----GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyVal 77  
Db 378 AGCAACTACAAAATCCCGTCGATCGCTCGGCAACGTGCATTCGGTGCTCA----- 428  
Qy 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 97  
Db 429 -----GCGGTGGCGGAGCAGTACATCGACCTG 455  
Qy 98 PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeu 117  
Db 456 GTGTCCACCGGTGCTCCGGGTAAATACTTCTCTCCCGGACAG----- 497  
Qy 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySer-----ProValPheAspLys 134

Db 498 -----ACCATCACCAGGACCGTTCCCGAGTGAGATCGGGCGCGCTGGACAAT 548  
Qy 135 AsnGlyArg-----LeuIleGlyLeuAlaPheAspGlyAsn 146  
Db 549 TCCAAATCCCGGTTGGCGCATTCGCCACGAGAGATCGCTTGCTGCTCGACGAGACC 608  
Qy 147 TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164  
Db 609 GCGCAAGCGTGGTGGG-----CTGGAGACCCCGCTTGCACCGTTGGTC 653

## RESULT 14

US-09-997-181-39  
; Sequence 39, Application US/09997181  
; Publication No. US20030049269A1  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: immunostimulatory Peptides  
; FILE REFERENCE: 61257  
; CURRENT APPLICATION NUMBER: US/09/997,181  
; PRIOR FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/447,135  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08/990,823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 39  
; LENGTH: 841  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(841)  
; OTHER INFORMATION: n is a, c, g, or t/u.  
US-09-997-181-39

Alignment Scores:  
Pred. No.: 0.491 Length: 841  
Score: 77.50 Matches: 40  
Percent Similarity: 32.91% Conservative: 12  
Best Local Similarity: 25.32% Mismatches: 45  
Query Match: 7.73% Indels: 61  
DB: 9 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x US-09-997-181-39 (1-841)

Qy 36 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArg---MetSerTyr 54  
Db 276 CTGATTCG-----ACGCCAATGTGACCTACCGGGGTATCACCATT 317  
Qy 55 GlySerIleLysGlyTyrGluProGlnAsp----- 64  
Db 318 GGCAAGGTTACTGCGTCGAGCCACCACCGAGGCGCCAGTGACGATGAGCATGCC 377  
Qy 65 -----GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyVal 77  
Db 378 AGCAACTACAAAATCCCGTCGATCGCTCGGCAACGTGCATTCGGTGCTCA----- 428  
Qy 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 97  
Db 429 -----GCGGTGGCGGAGCAGTACATCGACCTG 455  
Qy 98 PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeu 117  
Db 456 GTGTCCACCGGTGCTCCGGGTAAATACTTCTCTCCCGGACAG----- 497  
Qy 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySer-----ProValPheAspLys 134

```
Db 498 -----ACCATCAAGGCGACCGTTCCAGTGAGATCGGGCGCGCTGGACAAT 548
QY 135 AsnGlyArg-----LeuileGlyLeuAlaPheAspGlyAsn 146
Db 549 TCCAAATCGGGTTGGCGCATTCGCCACGAGAGATCGGCTTGCTCGCAGCAGACC 608
QY 147 TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164
Db 609 GCGCAAGCGGTGGGTGGG-----CTGGGACCCCGCTTGCAACGGTTGGTC 653
```

## RESULT 15

US-09-997-182-39

; Sequence 39, Application US/09997182

; Publication No. US20030049263A1

; GENERAL INFORMATION:

; APPLICANT: Nano, Francis

; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding

; TITLE OF INVENTION: immunostimulatory peptides

; FILE REFERENCE: 61258

; CURRENT APPLICATION NUMBER: US/09/997,182

; CURRENT FILING DATE: 2001-11-28

; PRIOR FILING DATE: 09/447,135

; PRIOR FILING DATE: 2000-01-03

; PRIOR APPLICATION NUMBER: 08/990,823

; PRIOR FILING DATE: 1997-12-15

; PRIOR APPLICATION NUMBER: US 96/10375

; PRIOR FILING DATE: 1996-06-14

; PRIOR APPLICATION NUMBER: 60/000,254

; PRIOR FILING DATE: 1995-06-15

; NUMBER OF SEQ ID NOS: 169

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 39

; LENGTH: 841

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)..(841)

; OTHER INFORMATION: n is a, c, g, or t/u.

US-09-997-182-39

## Alignment Scores:

Pred. No.:	0.491	Length:	841
Score:	77.50	Matches:	40
Percent Similarity:	32.91%	Conservative:	12
Best Local Similarity:	25.32%	Mismatches:	45
Query Match:	7.73%	Indels:	61
DB:	9	Gaps:	8

US-10-008-355-2\_copy\_522\_712 (1-191) x US-09-997-182-39 (1-841)

```
QY 36 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPhetThrMetArg---MetSerTyr 54
Db 276 CTGTATCCG-----ACGCCAAATGTGACCTACCGCGGTATCACCATT 317
QY 55 GlySerIleLysGlyTyrGluProGlnAsp----- 64
Db 318 GGCAAGGTACTGCGCTCGAGCCGCCACGAGGCGGCACGAGTGACGATGAGCATCGCC 377
QY 65 -----GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyVal 77
Db 378 AGCAACTACAAAATCCCGCTCGATGCTCGCGCAACGTCGATTCCGGTGCTCA----- 428
QY 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu 97
Db 429 -----CCGGTGGCGGAGAGTACATCGACCTG 455
QY 98 PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlnLeuHisIleAlaPheLeu 117
Db 456 GTGTCCACCGGTGCTCCGGGTAATACTTCTCTCCGACAG----- 497
QY 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySer-----ProValPheAspLys 134
```

```
Db 498 -----ACCATCAAGGCGACCGTTCCAGTGAGATCGGGCGCGCTGGACAAT 548
QY 135 AsnGlyArg-----LeuileGlyLeuAlaPheAspGlyAsn 146
Db 549 TCCAAATCGGGTTGGCGCATTCGCCACGAGAGATCGGCTTGCTCGCAGCAGACC 608
QY 147 TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164
Db 609 GCGCAAGCGGTGGGTGGG-----CTGGGACCCCGCTTGCAACGGTTGGTC 653
```

Search completed: May 23, 2003, 13:41:24

Job time : 55.1296 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 08:44:48 ; Search time 541.272 Seconds  
(without alignments)  
5714.937 Million cell updates/sec

Title: us-10-008-355-2\_COPY\_522\_712

Perfect score: 1003

Sequence: 1 SXSVTAAARAIQADAMANY.....LFMDKWGCPRLIQELKLI 191

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110401\_2551/app\_query.fasta\_1.1230  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355.@CGN\_1.1.2013.@runat\_16052003\_110401\_2551 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	314	31.3	426	17	BH400391 AG-ND-147
2	136	13.6	822	17	BH371846 AG-ND-162
3	90.5	9.0	1102	13	BM562210 AGENCOURT
4	89.5	8.9	630	12	BG491629 602535962
5	89.5	8.9	667	13	BI391728 ppln.pk0
6	89.5	8.9	701	10	BE268652 601125126
7	89.5	8.9	710	14	BM38106 K-EST0114
8	89.5	8.9	852	14	BQ215538 AGENCOURT
9	89.5	8.9	853	12	BE799506 601589315
10	89.5	8.9	881	13	BM456284 AGENCOURT
11	89.5	8.9	899	13	BM449938 AGENCOURT
12	89.5	8.9	904	14	BQ645307 AGENCOURT
13	89.5	8.9	942	14	BQ652808 AGENCOURT
14	89.5	8.9	943	12	BG174574 602334417
15	89.5	8.9	960	14	BQ646879 AGENCOURT
16	89.5	8.9	982	14	BQ645695 AGENCOURT
17	89.5	8.9	1004	14	BQ960408 AGENCOURT
18	89.5	8.9	1015	14	BQ935052 AGENCOURT
19	89.5	8.9	1027	14	BM928590 AGENCOURT
20	89.5	8.9	1031	13	BM550804 AGENCOURT
21	89.5	8.9	1036	14	BQ957384 AGENCOURT
22	89.5	8.9	1039	14	BM911788 AGENCOURT
23	89.5	8.9	1045	14	BM914762 AGENCOURT
24	89.5	8.9	1052	14	BM809287 AGENCOURT
25	89.5	8.9	1074	14	BM909854 AGENCOURT
26	89.5	8.9	1092	14	BM908687 AGENCOURT
27	89.5	8.9	1097	13	BM455151 AGENCOURT
28	89.5	8.9	3022	11	BC014232 Homo sapi
29	88.5	8.8	656	13	BG918792 AGENCOURT
30	86.5	8.6	947	14	BQ649391 AGENCOURT
31	85.5	8.5	759	10	BE535285 601058826
32	85.5	8.5	1062	13	BM464181 AGENCOURT
33	84.5	8.4	428	13	BI080699 602878440
34	83.5	8.3	412	12	BF841384 RC2-HT107
35	83.5	8.3	694	9	AU131725
36	83.5	8.3	743	9	AU134050
37	83.5	8.3	861	13	BI116059 602866455
38	83.5	8.3	960	14	BQ892534 AGENCOURT
C 39	83	8.3	571	12	BG602643 EST501733
40	83	8.3	655	13	BI334196 602997681
41	83	8.3	1385	14	BM911372 AGENCOURT
42	82.5	8.2	412	10	BB674829 BB674829
43	82.5	8.2	491	9	AL034929 m8708a54
C 44	82.5	8.2	558	13	BM204311 C0267F10-
45	82	8.2	893	12	BF699325 602125888

ALIGNMENTS

RESULT 1  
BH400391/c  
LOCUS  
DEFINITION BH400391 AG-ND-147H4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4,  
DNA sequence.  
ACCESSION BH400391 GI:17346607  
KEYWORDS GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoides;  
Anopheles.  
REFERENCE 1 (bases 1 to 426)  
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: AG-ND-147H4.TR  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@tigr.org  
 This clone is from an A. gambiae BAC library (ND-TAM) provided by  
 F.H. Collins and sequenced by The Institute for Genomic Research  
 (TIGR). The BAC library was generated from A. gambiae PEST strain  
 DNA. All DNA was extracted from newly hatched first instar larvae  
 to minimize the inclusion of DNA from microorganisms that inhabit  
 the gut. The DNA is derived from mixed sexes of larvae. The BAC  
 library was constructed at Texas A&M University BAC Center  
 University, College Station, Texas 77843-2123, USA using a HindIII  
 partial digest.  
 Seq primer: M13 For  
 Class: BAC ends.  
 Location/Qualifiers  
 1..426  
 /organism="Anopheles gambiae"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-147H4"  
 /clone\_lib="ND-TAM"  
 /note="Vector: pECBAC1; Site\_1: HindIII"  
 111 a 79 c 76 g 160 t  
 BASE COUNT  
 ORIGIN

Alignment Scores:	3.63e-32	Length:	426
Pred. No.:	314.00	Matches:	65
Score:	65.41%	Conservative:	22
Percent Similarity:	48.87%	Mismatches:	40
Best Local Similarity:	48.87%	Indels:	6
Query Match:	31.31%	Gaps:	2
DB:	17		
US-10-008-355-2_COPY_522_712 (1-191) x BH400391 (1-426)			
Qy	23	GlulysgLyLysArgLeuPhePheAlaGlyLeuArgGluuMetTyrProGlyArgAlaLeu	42
Db	399	CAAAAACACCCGGTTTATTCATGGCGGAAATNATAAATCTATGCCTGAAAAAATTC	340
Qy	43	ProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIle	57
Db	339	TATCCGGATGCTAACTCTACGATCAGGTTAACTTATGGTCAGATCGAGCGCTTCTCTAA	280
Qy	58	LySGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyVal	77
Db	279	AGAGCAGACAGAGACTACAGGTCATCAAGCAGAATTACTATACCAATGGAAGGTATG	220
Qy	78	LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu	97
Db	219	ATTAAGAAGTACAAAGAAGGTGACGAAGATTCGATCTTCCACAGGAGCTCTTGATCTT	160
Qy	98	PheArgThrLysAsnTyrGlyArgTyrAlaGlu--AsnGlyGlnLeuHisIleAlaPhe	116
Db	159	TATAAAGAAGAAATTTATGGTATCTACAAGACAAAGACGGGCAACTTCATGTAAACTTC	100
Qy	117	LeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGly	136
Db	99	CTTTCATAACGATATTACAGGAGGTAACTCAGGTCTCTCCAATTTATCGATGCTTACGCA	40
Qy	137	ArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAla	149
Db	39	AGACTTATAGGCTCTGCATTTGACGGAAACAGTGAAGCT	1

RESULT 2					
BH371846					
LOCUS	BH371846	822 bp	DNA	linear	GSS 10-DEC-2001

DEFINITION	AG-ND-162M17.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162M17
ACCESSION	, DNA sequence.
VERSION	BH371846
KEYWORDS	BH371846.1 GI:17317971
SOURCE	GSS.
ORGANISM	African malaria mosquito.
	Anopheles gambiae
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
	Anopheles.
REFERENCE	1 (bases 1 to 822)
AUTHORS	Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
TITLE	Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL	Unpublished (2001)
COMMENT	Other_GSSs: AG-ND-162M17.TR
	Contact: Brendan J Loftus
	Department of Eukaryotic Genomics
	The Institute for Genomic Research
	9712 Medical Center Dr., Rockville, MD 20850, USA
	Tel: 301 838 0208
	Fax: 301 838 3543
	Email: bjlftus@tigr.org

This clone is from an *A. gambiae* BAC library (NU-1AM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For  
Class: BAC ends.

<b>FEATURES</b>	<b>SOURCE</b>
-----------------	---------------

```
1. 7021
/organism="Anopheles gambiae"
/strain="pEST"
/db_xref="taxon:7165"
/clone="AG-ND-162M17"
/clone_lib="ND-TAN"
/notes="vector: pECBAC1; Site:1: HindIII"
235 a 142 c 155 q 290 t
BASE COUNT
```

```

BASE COUNT      235 a      142 c      155 g      290 t
ORIGIN

Alignment Scores:
Pred. No. :      1.68e-07      Length:      822
Score:      136.00      Matches:      27
Percent Similarity:      79.07%      Conservatives:      7
Best Local Similarity:      62.79%      Mismatches:      9
Query Match:      13.56%      Indels:      0
DB:      17      Gaps:      0

```

US-10-008-355-2\_COPY\_522\_712 (1-191) x BH371846 (1-822)

Qy 149 AlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIle 168  
 Db 2 GCCTTTAAGCGGTGACATGTGTTTGAACCTTAATTACAAGAAGACGATTAACTAGACGTT 61

Dd	2	GCCTTAAGCGGTGACATTGTTTGTGTAACCTTAAATACAAAGAACGATTACCGTAGACGTT	61
Qy	169	ArgTyrValLeuPheMetIleAspLysTrpGlyClnCysProArgLeuIleGlnGluLeu	188
Dd	62	AGATACCTACTCTTGGTGGTAATTCGACACCTGCTTAAAGCTTAATTAATACGGAATTCG	121

Qy 189 LysLeuIle 191

Db 122 ACTTTAGTA 130

### RESULT 3

BM307210  
LOCUS

### DEFINITION

ACCESSION  
NUMBER

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

BM5022410	BM562210	1102 bp	mRNA	linear	EST 20-FEB-2002
LOCUS					
DEFINITION	AGENCOURT	6597467	NIH MGC 41	Homo sapiens	cdna clone IMAGE:5480814

5', mRNA sequence.

BM362210	BM562210.1	GT:18808104
ACCESSION	VERSION	



```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1102)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2004 row: e column: 07
High quality sequence stop: 448.
BASE COUNT 301 a 272 c 304 g 215 t 10 others
ORIGIN
1..1102
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5480814"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."
BASE COUNT 301 a 272 c 304 g 215 t 10 others
ORIGIN
1..1102
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5480814"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

FEATURES
source
Location/Qualifiers
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4684958"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

BASE COUNT 145 a 155 c 177 g 153 t
ORIGIN
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4684958"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

Alignment Scores:
Pred. No.: 0.225 Length: 630
Score: 89.50 Matches: 42
Percent Similarity: 38.32% Conservative: 22
Best Local Similarity: 25.15% Mismatches: 56
Query Match: 8.92% Indels: 47
DB: 12 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x BM562210 (1-1102)
Qy 19 AlaTyrAlaIleGluLysGlyLeuArgLeuPheAlaGlyLeuArgGluMetTyrPro 38
Db 146 AGCTATGGGGTCAGAAAGGGCGGTATGCTTCGAGATGAAGATCAATGAGGAATCTCC 205
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52
Db 206 GTGAAGCACCTTCGCTACAGAGCGCTGACCCACGCGTGGTCCGATCGCTGGTCCCTG 265
Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHis 71
Db 266 GACTCTCGACGACCCAGCTAGGCGAAGAGCGCTTCTCC-----TATGGCTATGGA 316
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
Db 317 CGCACTGGGAAG-----AAGTCCACCAAT----- 340
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110
Db 341 -----AGCGGTTTGAAAACTACCGAGACAAGTTTGCAGAG----- 376
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly----- 124
Db 377 -----AACGATGTGATTGGCTGCTTTGCCGATTGTTGAA 409

```

```

Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 410 TGTGGAATGAGTGGAACTGCTTTTACCAAGAATGGAAGTGGATTGCTTTC 469
Qy 144 AspGlyAsnTyrGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 470 CGAATCCAGAGAGAGCCCTTGGGGGTGAGCCCTCTATCTCTCATGTCTGTGGAAGA 529
Qy 164 IleSerValAspIleArgTyr 170
Db 530 TGGCAGTGGAGCTCGACTTC 550

RESULT 4
BG491629 630 bp mRNA linear EST 27-MAR-2001
LOCUS 602535962F1 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:4684958 5',
DEFINITION mRNA sequence.
ACCESSION BG491629
VERSION BG491629.1 GI:13453141
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 630)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1493 row: 1 column: 15
High quality sequence stop: 630.
FEATURES
source
Location/Qualifiers
1..630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4684958"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."

```

```

Db 207 GTGAGCAGCTTCGCTACAGAGCCTGACCCCGACGTGGTCCGTATCGCGCTGTCCTG 266
      ... ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHis 71
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 267 GACTCCTGCACACCCAGCTAGGCGAAGAGCCTTTCCTCC-----TATGGCTATGGA 317
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 318 GGCACCTGGAG-----AGTCCACCAAT----- 341
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 342 -----AGCCGGTTTGAACATACGGAGACAAGTTTGCAGAG----- 377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly----- 124
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 378 -----AACGATGTGATGGTCTGCTTTGCGGATTTTGAA 410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 411 TGTGGAATGACGTGGAACCTCTTTTACCAGAAATGGAAGTGGATGGGCGGATTC 470
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 471 CGAATCCAGAAAGGAGCCTTGGGGGTCAGGCCCTCTATCTCTCATGCTCGTGAAGAA 530
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 164 IleSerValAspIleArgTyr 170
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 531 TGGCAGTGGAGTCAACTTC 551
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BI391728 667 bp mRNA linear EST 06-AUG-2001
LOCUS
DEFINITION popln.pk005.f8 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library Gallus gallus cDNA clone pgpin.pk005.f8 5' similar to
gb|AAH04057.1|AAH04057 (BC004057) Similar to transforming, acidic
colled-coil containing protein 2 [Mus musculus], mRNA sequence.
ACCESSION BI391728
VERSION BI391728.1 GI:15085010
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 667)
Porter, T.E. and Cogburn, L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library USDA/IFAFS Animal Genome Project
JOURNAL Unpublished (2001)
COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.

FEATURES
source
1..667
/organism="Gallus gallus"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="pgpin.pk005.f8"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library"
/sex="Male and female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5
,w7,w9)"
/lab_host="E. Coli EMDH108"
/note="vector: pcWSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"

```

```

BASE COUNT 223 a 116 c 178 g 132 t 18 others
ORIGIN
Alignment Scores:
Pred. NO.: 0.247 Length: 667
Score: 89.50 Matches: 28
Percent Similarity: 47.86% Conservative: 28
Best Local Similarity: 23.93% Mismatches: 54
Query Match: 8.92% Indels: 7
DB: 13 Gaps: 2

US-10-008-355-2_copy_522_712 (1-191) x BI391728 (1-667)
Qy 6 AlaAlaAlaAAGAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGly 25
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 GCTGCACTACCACTGCCAGACAGAGATTGTGGCCAGGAAAGAGAGATGCTCGAGTGG 235
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 26 LysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAsp 45
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 AAGAGAAATATGAAGAAAGCAGAGGCAAGTGTGAAATGAGGAAATAGTTTCAGAA 295
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 46 AlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGly 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 296 TATGAGAAGACGATTTGCTCAGATGATAGGC-----AAGCCCGAGGATGACACAGAG 349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 66 AlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSer 85
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 350 AAGTCTGTCTCCATCACACTGTTTCAGCAGCTGATCGTGGAGAGAGCAGCGCTGCGG 409
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 86 AspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArg 105
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 GATCTGAACCTCAGTGGAGAAATCTCTGCGACATCTTTTCAGG-----AGA 454
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 106 TyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAspIle 122
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 TATCAAAAATCAAGGAAGTGTGTGGAGGGTTCGGAAGAACGAAGAAGTA 505
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
BI268652 701 bp mRNA linear EST 13-JUL-2000
LOCUS
DEFINITION 601125126F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:3344958 5',
mRNA sequence.
ACCESSION BI268652
VERSION BI268652.1 GI:9142260
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 701)
NTH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI32 row: c column: 07
High quality sequence start: 8
High quality sequence stop: 678.
FEATURES
Location/Qualifiers
1..701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3344958"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: xhoI; Site_2:

```

ECORI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).

BASE COUNT 161 a 170 c 210 g 160 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.269 Length: 701  
Score: 89.50 Matches: 42  
Percent Similarity: 38.32% Conservativeness: 22  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 8.92% Indels: 47  
DB: 10 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BE268652 (1-701)

Qy 19 AlaTyrAlaIleGluLysGlyLeuPhePheAlaGlyLeuArgGluMetTyrPro 38  
Db 134 AGCTATGGGTCAGAAAGGGCGGTATGCTTCGAGATGAAGATCAATGAGAAATCTCC 193  
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52  
Db 194 GTGAAGCACCTTCGCTACAGACCTGACCCCGGTCGTCGATGCTGCTGCTGCTGCTG 253  
Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHis 71  
Db 254 GACTCTGCACACCCAGCTAGGCGAGAGCCTTTCTCC-----TATGGCTATGGA 304  
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91  
Db 305 GGCACCTGGGAAG-----AAGTCCACCAAT----- 328  
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly----ArgTyrAlaGluAsnGly 110  
Db 329 -----AGCGGTTTGAAGAACTACGAGACAAGTTTGCAGAG----- 364  
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly----- 124  
Db 365 -----AAGCATGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 397  
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143  
Db 398 TGTGGAATGACGTGGAACTGCTTTTACCAAGAATGAAAGTGGATGGCGCATTTCTTC 457  
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163  
Db 458 CGAATCCAGAAGGAGCCTGGGGGTCAGGCCCTCTATCTCTGCTGCTGCTGCTGCTG 517  
Qy 164 IleSerValAspIleArgTyr 170  
Db 518 TCCGAGTGGAGTCAACTTC 538

RESULT 7  
BM838106 710 bp mRNA linear EST 06-MAR-2002  
LOCUS  
DEFINITION K-EST0114293 S9SNU601 Homo sapiens cDNA clone S9SNU601-73-C05 5',  
mRNA sequence.

ACCESSION BM838106  
VERSION BM838106.1 GI:19194515  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 710)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.

TITLE 21c Frontier Korean EST Project 2001

# JOURNAL COMMENT

Unpublished (2002)

Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr

High quality sequence stop: 710.  
Plate: 73 row: C column: 05

## FEATURES

source  
1..710  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S9SNU601-73-C05"  
/clone\_lib="S9SNU601"  
/sex="M"  
/tissue\_type="Ascites"  
/cell\_type="Epithelial"  
/lab\_host="Top10F"  
/note="Organ: Stomach; Vector: pME18-FL3; Site\_1: XhoI;  
Site\_2: XhoI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including SfiI  
site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized with Superscript II using SfiI  
oligo-dT primer. After first strand synthesis, RNA was  
degraded by NaOH treatment and cDNA was amplified by PCR  
reaction. The PCR products were digested with SfiI and  
cloned into DraIII- digested pME18S-FL3 vector. The  
obtained cDNA vectors were used for transfection of  
competent cells E. coli Top10F by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

BASE COUNT 163 a 182 c 201 g 164 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0.275 Length: 710  
Score: 89.50 Matches: 42  
Percent Similarity: 38.32% Conservativeness: 22  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 8.92% Indels: 47  
DB: 14 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BM838106 (1-710)

Qy 19 AlaTyrAlaIleGluLysGlyLeuPhePheAlaGlyLeuArgGluMetTyrPro 38  
Db 115 AGCTATGGGTCAGAAAGGGCGGTATGCTTCGAGATGAAGATCAATGAGAAATCTCC 174  
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52  
Db 175 GTGAAGCACCTTCGCTACAGACCTGACCCCGGTCGTCGATGCTGCTGCTGCTGCTG 234  
Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHis 71  
Db 235 GACTCTGCACACCCAGCTAGGCGAGAGCCTTTCTCC-----TATGGCTATGGA 285  
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91  
Db 286 GGCACCTGGGAAG-----AAGTCCACCAAT----- 309  
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly----ArgTyrAlaGluAsnGly 110  
Db 310 -----AGCGGTTTGAAGAACTACGAGACAAGTTTGCAGAG----- 345  
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly----- 124  
Db 346 -----AAGCATGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 378

```

Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 379 TGTGAAATGACGTGGAACCTGCTCTTTTACCAAGAATGAAAGTGGATGGCATTGCTTTC 438
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 439 CGAATCAGAGGAAGCCTTGGGGGGTGCAGGCCCTCTATCCTCATGCTCGTGGTGAAGAAT 498
Qy 164 IleSerValAspIleArgTyr 170
Db 499 TGGCAGTGGAGTCAACTTC 519

RESULT 8
LOCUS BQ215538 852 bp mRNA linear EST 02-MAY-2002
DEFINITION AGENCOURT_7566326 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6043333
5' mRNA sequence.
ACCESSION BQ215538
VERSION BQ215538.1 GI:20396938
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13284 row: c column: 14
High-quality sequence stop: 673.
Location/Qualifiers
1..852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6043333"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 198 a 228 c 239 g 186 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.372 Length: 852
Score: 89.50 Matches: 42
Percent Similarity: 38.32% Conservative: 22
Best Local Similarity: 25.15% Mismatches: 56
Query Match: 8.92% Indels: 47
DB: 14 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x BQ215538 (1-852)

Qy 19 AlaTyrAlaIleGluLysGlyArgLeuPheAlaGlyLeuArgGluMetTyrPro 38
Db 103 AGCTATGGGTGCAGAGGGCCGTGTATGCTTCGAGATGAAGATCAATGAGAAATCTCC 162
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52
Db 163 GTGAAGCACCCTTCGCTATACAGACGCTGACCCACCACCGTTCGCTATGCTGCTCCTG 222

```

```

Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHis 71
Db 223 GACTTCGTCAGCACCCAGCTAGGCGAAGAGCGCTTTCTCC-----TATGGCTATGGA 273
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
Db 274 GGCACCTGGGAAG-----AAGTCCACCAAT----- 297
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110
Db 298 -----AGCCGGTTTGAANAACACGAGACACAGTTTGCAGAG----- 333
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly----- 124
Db 334 -----AAGCATCTGATTGGCTGCTTTTGCAGATTTTGA 366
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 367 TGTGAAATGACGTGGAACCTGCTCTTTTACCAAGAATGAAAGTGGATGGCATTGCTTTC 426
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 427 CGAATCAGAGGAAGCCTTGGGGGGTGCAGGCCCTCTATCCTCATGCTGCTGCTGAAGAAT 486
Qy 164 IleSerValAspIleArgTyr 170
Db 487 TGGCAGTGGAGTCAACTTC 507

RESULT 9
LOCUS BE799506 853 bp mRNA linear EST 20-SEP-2000
DEFINITION 601589315F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943490 5'
mRNA sequence.
ACCESSION BE799506
VERSION BE799506.1 GI:10220704
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 853)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM798 row: n column: 03
High quality sequence start: 24
High quality sequence stop: 790.
Location/Qualifiers
1..853
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3943490"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 202 a 203 c 255 g 193 t

```

## ORIGIN

## Alignment Scores:

Pred. No.: 0.373 Length: 853  
Score: 89.50 Matches: 42  
Percent Similarity: 38.32% Conservative: 22  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 8.92% Indels: 47  
DB: 12 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BE799506 (1-853)

Qy 19 AlaTyrAlaIleGluLysGlyLeuPheAlaGlyLeuArgGluMetTyrPro 38  
Db 288 AGCTATGGGTGTCAGAGGGCGGTATGCTTCGAGATGAAGATCAATGAGCAATCTCC 347  
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPhetThrMetArgMet----- 52  
Db 348 GTGAAGCACCTTCCTGCTACAGACCTGACCCCGTGGTCTGCTATGCGGTGCTCCCTG 407  
Qy 53 ---SertYrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHis 71  
Db 408 GACTCTCCACACCCAGCTAGCGAAGAGCCTTCTCC-----TATGGCTATGGA 458  
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91  
Db 459 GGCACCTGGGAAG-----AAGTCCACCAAT----- 482  
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110  
Db 483 -----AGCCGGTTTGAAACTACGAGACAAAGTTGCAAG----- 518  
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly-----AAGTCCACCAAT----- 124  
Db 519 -----AAGTCCACCAAT----- 124  
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143  
Db 552 TGTGGAATACGCTGGAAGTCTCTTTTACCAAGAAATGAAAGTGGATGGCTTTC 611  
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGluArgThr 163  
Db 612 CGATCCAGAGGAGGCTTGGGGGTGAGCCCTCTATCTCTATCTCTGCTGCTGGAAGAA 671  
Qy 164 IleSerValAspIleArgTyr 170  
Db 672 TGCAGTGGAGTTCAACTTC 692

## RESULT 10

BM456284

LOCUS AGENCOURT\_6409930 NIH\_MGC\_85 Homo sapiens cdna clone IMAGE:5496635

DEFINITION 5', mRNA sequence.

ACCESSION BM456284

VERSION BM456284.1 GI:18505324

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 881)

NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-f@mail.nih.gov

Tissue Procurement: Lou Staudt

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12125 row: h column: 12

High quality sequence stop: 664.

## FEATURES

## source

1..881  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5496635"  
/clone\_lib="NIH\_MGC\_85"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.867 Kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC Library."

BASE COUNT 207 a 224 c 256 g 192 t

## ORIGIN

## Alignment Scores:

Pred. No.: 0.393 Length: 881  
Score: 89.50 Matches: 42  
Percent Similarity: 38.32% Conservative: 22  
Best Local Similarity: 25.15% Mismatches: 56  
Query Match: 8.92% Indels: 47  
DB: 13 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BM456284 (1-881)

Qy 19 AlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrPro 38  
Db 273 AGCTATGGGTGTCAGAGGGCGGTATGCTTCGAGATGAAGATCAATGAGCAATCTCC 332  
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPhetThrMetArgMet----- 52  
Db 333 GTGAAGCACCTTCCTGCTACAGACCTGACCCCGTGGTCTGCTATGCGGTGCTCCCTG 392  
Qy 53 ---SertYrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHis 71  
Db 393 GACTCTCCACACCCAGCTAGCGAAGAGCCTTCTCC-----TATGGCTATGGA 443  
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91  
Db 444 GGCACCTGGGAAG-----AAGTCCACCAAT----- 467  
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110  
Db 468 -----AGCCGGTTTGAAACTACGAGACAAAGTTGCAAG----- 503  
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly----- 124  
Db 504 -----AAGTCCACCAAT----- 467  
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143  
Db 537 TGTGGAATGACGTGGAAGTCTCTTTTACCAAGAAATGAAAGTGGATGGCTTTC 596  
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGluArgThr 163  
Db 597 CGATCCAGAGGAGGCTTGGGGGTGAGCCCTCTATCTCTATCTCTGCTGCTGGAAGAA 656  
Qy 164 IleSerValAspIleArgTyr 170  
Db 657 TGCAGTGGAGTTCAACTTC 677  
RESULT 11  
BM449938  
LOCUS AGENCOURT\_6392962 NIH\_MGC\_72 Homo sapiens cdna clone IMAGE:5528678  
DEFINITION 5', mRNA sequence.  
ACCESSION BM449938  
VERSION BM449938.1 GI:18498978  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 899)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC/DCTD/DPMP  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM12205 row: o column: 15  
 High quality sequence stop: 670.

FEATURES  
 source  
 1. 899  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5528678"  
 /clone\_lib="NIH\_MGC\_72"  
 /tissue\_type="melanotic melanoma"  
 /note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. primer: oligo dt.  
 Average insert size 2 kb. Library constructed by Life  
 Technologies."

BASE COUNT 209 a 244 c 250 g 194 t 2 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.407 Length: 899  
 Score: 89.50 Matches: 42  
 Percent Similarity: 38.32% Conservative: 22  
 Best Local Similarity: 25.15% Mismatches: 56  
 Query Match: 8.92% Indels: 47  
 DB: 13 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BM449938 (1-899)

Qy 19 AlaTyrAlaIleGluLysGlyValLeuArgGluMetTyrPro 38  
 Db 102 AGCTATGGGTCACAGGGCGCGTATGCTTCGAGATGAAGATCAATGAGAAATCTCC 161  
 Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52  
 Db 162 GTGAAGCACCTTCGCTCTACAGAGCCCTGACCCCGCTGGTCCGTATCGGCTGCCCTG 221  
 Qy 53 ---SertYrGlySerIleLysGlyValLeuProGlnAspGlyAlaTrpTyrAsnTyrHis 71  
 Db 222 GACTCTGCACACCCAGCTAGGCGAGAGCCCTTCTCC-----TATGGTATGGA 272  
 Qy 72 ThrThrGlyTysGlyValLeuLysGlnAspProLysSerAspGluPheAlaValGln 91  
 Db 273 GGCACCTGGGAAG-----AAGTCCACCAAT----- 296  
 Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly----ArgTyrAlaGluAsnGly 110  
 Db 297 -----AGCCGGTTTGAACACTACGAGACAAGTTTGCAGAG----- 332  
 Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly----- 124  
 Db 333 -----AACGATGTGATGGTCTCTTTCGCGGATTTGAA 365  
 Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143  
 Db 366 TGTGCAAAATGACGTGGAACTGCTCTTTTACCAAGAATGAAAGTGGATGGCGCATTC 425  
 Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163  
 Db 144 ----- 163

Db 426 CGAATCCAGAAGAGCCTTGGGGGGTTCAGCCCTCTATCTCATGTCTCTGGTGAAGAAT 485  
 Qy 164 IleSerValAspIleArgTyr 170  
 Db 486 TGGCGAGTGGAGTTCAACTTC 506

RESULT 12  
 BQ645307 904 bp mRNA linear EST 15-JUL-2002  
 LOCUS AGENCOURT\_8302363 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6271209  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BQ645307  
 VERSION BQ645307.1 GI:21769479  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 904)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLCM2447 row: j column: 10  
 High quality sequence stop: 741.

FEATURES  
 source  
 1. 904  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6271209"  
 /clone\_lib="NIH\_MGC\_100"  
 /tissue\_type="hepatocellular carcinoma, cell line"  
 /note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 217 a 227 c 259 g 201 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.411 Length: 904  
 Score: 89.50 Matches: 42  
 Percent Similarity: 38.32% Conservative: 22  
 Best Local Similarity: 25.15% Mismatches: 56  
 Query Match: 8.92% Indels: 47  
 DB: 14 Gaps: 8

US-10-008-355-2\_COPY\_522\_712 (1-191) x BQ645307 (1-904)

Qy 19 AlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrPro 38  
 Db 239 AGCTATGGGTCACAGGGCGCGTATGCTTCGAGATGAAGATCAATGAGAAATCTCC 298  
 Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52  
 Db 299 GTGAGCACCTTCGCTCTACAGAGCCCTGACCCCGCTGGTCCGTATCGGCTGCCCTG 358  
 Qy 53 ---SertYrGlySerIleLysGlyValLeuProGlnAspGlyAlaTrpTyrAsnTyrHis 71  
 Db 53 ----- 71

```

Db 359 GACTCTCTGCAGCACCCAGCTAGCGAAGAGCCTTTCTCC-----TATGGCTATGGA 409
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
Db 410 GGCACCTGGGAG-----AAGTCCACCAAT-----433
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTrpGly---ArgTyrAlaGluAsnGly 110
Db 434 -----AGCCGGTTTGAAAACTACGAGACAAGTTTGCAGAG-----469
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly-----124
Db 470 -----AAGATGTGATGGCTCTCTTGGCGATTGAA 502
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 503 TGTGGAATGACGTGGAACTCTCTTTACCAAGAATGGAAGTGGATGGCGATTGCTTTC 562
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 563 CGAATCCAGAGGAGGCTTGGGGGTTCAGGCCCTCTATCTCTCATCTGCTGGTGAAGAAT 622
Qy 164 IleSerValAspIleArgTyr 170
Db 623 TGGCAGTGGAGTTCAACTTC 643

RESULT 13
LOCUS BQ652808 942 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8490114 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6298491
5', mRNA sequence.
ACCESSION BQ652808
VERSION BQ652808.1 GI:21776980
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 942)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2509 row: k column: 04
High quality sequence stop: 621.
Location/Qualifiers
1. 942
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6298491"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected by ligation for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 222 a 240 c 275 g 205 t
ORIGIN

```

```

Alignment Scores: 0.44 Length: 942
Pred. No.: 89.50 Matches: 42
Percent Similarity: 38.32% Conservative: 22
Best Local Similarity: 25.15% Mismatches: 56
Query Match: 8.92% Indels: 47
DB: 14 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x BQ652808 (1-942)
Qy 19 AlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPro 38
Db 239 AGCTATGGGTTCAGAGGGGCGGTGTATCTTGAGATGAAGATCAATGAGGAATCTCC 298
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet-----52
Db 299 GTGAAGCACCTTCGCTCTACAGACGCTGACCCACGCTGGTCCGTATCGGCTGGTCCCTG 358
Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHis 71
Db 359 GACTCTCTGCAGCACCCAGCTAGCGAAGAGCCTTTCTCC-----TATGGCTATGGA 409
Qy 72 ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
Db 410 GGCACCTGGGAG-----AAGTCCACCAAT-----433
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTrpGly---ArgTyrAlaGluAsnGly 110
Db 434 -----AGCCGGTTTGAAAACTACGAGACAAGTTTGCAGAG-----469
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly-----124
Db 470 -----AAGCATGTGATGGCTCTCTTGGCGATTGAA 502
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 503 TGTGGAATGACGTGGAACTCTCTTTACCAAGAATGGAAGTGGATGGCGATTGCTTTC 562
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 563 CGAATCCAGAGGAGGCTTGGGGGTTCAGGCCCTCTATCTCTCATCTGCTGGTGAAGAAT 622
Qy 164 IleSerValAspIleArgTyr 170
Db 623 TGGCAGTGGAGTTCAACTTC 643

RESULT 14
LOCUS BQ174574 943 bp mRNA linear EST 06-FEB-2001
DEFINITION 602334417F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4457381 5',
mRNA sequence.
ACCESSION BQ174574
VERSION BQ174574.1 GI:12681277
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 943)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Glibert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10254 row: b column: 06
High quality sequence stop: 656.

```





Db 503 TGTGGAATGACGTGGAACTCTCTTTACCAAGAAATGAAAGTGGATGGGCATTGCTTTTC 562  
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163  
Db 563 CGAATCCAGAGGAGCCCTTGGGGGTCAGGCCCTCTATCCTCATGTCTGCTGAAGAAAT 622  
Qy 164 IleSerValAspIleArgTyr 170  
Db 623 TGCAGTGGAGTTCAACTTC 643

Search completed: May 23, 2003, 12:11:28  
Job time : 547.272 secs



GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 13:47:08 ; Search time 17.3636 Seconds  
(without alignments)  
1060.959 Million cell updates/sec

Title: US-10-008-355-2\_COPY\_522\_712

Perfect score: 1003

Sequence: 1 SKSVIAAARAIOADAMANAY.....LFMDKMGQCPRLIQELKLI 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 362588

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published\_Applications\_AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	1003	100.0	712	9	US-10-008-355-2
2	936	93.3	699	9	US-10-008-355-8
3	427.5	42.6	720	9	US-10-008-355-9
4	334	33.3	716	9	US-10-008-355-7
5	274	27.3	52	9	US-10-008-355-3
6	262.5	26.2	732	9	US-10-008-355-6
7	255.5	25.5	734	9	US-10-008-355-5
8	81.5	8.1	254	10	US-09-815-242-5351
9	81.5	8.1	254	10	US-09-815-242-12277
10	81	8.1	283	9	US-10-062-051-4
11	81	8.1	283	12	US-10-062-624-4
12	78	7.8	52	9	US-10-008-355-4
13	77.5	7.7	165	9	US-09-981-286A-33
14	77	7.7	332	10	US-09-767-041-21
15	76	7.6	889	9	US-09-909-567B-44
16	74.5	7.4	342	10	US-09-991-258-10
17	73.5	7.3	157	9	US-09-981-286A-1
18	73.5	7.3	169	9	US-09-981-286A-7
19	73	7.3	278	9	US-09-738-626-4716

20	73	7.3	336	9	US-09-987-107-44	Sequence 44, Appl
21	73	7.3	436	10	US-09-815-242-11154	Sequence 11154, A
22	73	7.3	1057	9	US-10-135-322-5	Sequence 5, Appli
23	73	7.3	1057	9	US-10-135-322-24	Sequence 24, Appl
24	73	7.3	1057	9	US-09-918-508-6	Sequence 6, Appli
25	72.5	7.2	280	12	US-10-062-994-4	Sequence 4, Appli
26	72.5	7.2	280	12	US-10-062-994-4	Sequence 4, Appli
27	72.5	7.2	337	9	US-09-987-107-46	Sequence 46, Appl
28	72.5	7.2	711	9	US-10-043-344-98	Sequence 98, Appl
29	72	7.2	283	12	US-10-059-964-40	Sequence 40, Appl
30	72	7.2	524	10	US-09-925-300-1538	Sequence 1538, Ap
31	71.5	7.1	165	9	US-09-981-286A-34	Sequence 34, Appl
32	71.5	7.1	289	9	US-10-135-322-10	Sequence 10, Appl
33	71.5	7.1	397	9	US-09-738-626-3831	Sequence 3831, Ap
34	71	7.1	323	9	US-10-101-464A-764	Sequence 764, App
35	71	7.1	804	9	US-10-101-464A-890	Sequence 890, App
36	70.5	7.0	568	9	US-09-418-176-4	Sequence 4, Appli
37	70	7.0	567	10	US-09-815-242-13511	Sequence 13511, A
38	70	7.0	1356	10	US-09-757-781-2	Sequence 2, Appli
39	69.5	6.9	341	10	US-09-815-242-11798	Sequence 11798, A
40	69.5	6.9	953	9	US-09-953-280-46	Sequence 46, Appl
41	69	6.9	432	10	US-09-815-242-10274	Sequence 10274, A
42	69	6.9	3169	9	US-10-114-170-257	Sequence 257, App
43	68.5	6.8	958	9	US-09-953-280-43	Sequence 43, Appl
44	68.5	6.8	1242	9	US-09-925-299-911	Sequence 911, App
45	68.5	6.8	1242	10	US-09-925-299-911	Sequence 911, App

ALIGNMENTS

RESULT 1  
US-10-008-355-2  
; Sequence 2, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 712  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-2

Query Match 100.0%; Score 1003; DB 9; Length 712;  
Best Local Similarity 100.0%; Pred. No. 1.2e-95;  
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	SKSVIAAARAIOADAMANAYATEKGRKLFAGLRREMPGRALPSDANFTMRMSYGSIKGY	60
DB	522	SKSVIAAARAIOADAMANAYATEKGRKLFAGLRREMPGRALPSDANFTMRMSYGSIKGY	581
QY	61	EPQDGAWYNYHTGKVLKQDPKSDFAVQENILDLFTTKNYGRYAENGQLHIAFLSN	120
DB	582	EPQDGAWYNYHTGKVLKQDPKSDFAVQENILDLFTTKNYGRYAENGQLHIAFLSN	641
QY	121	DITGGNSGSPFDKNGRLGLAFDGNWEAMSGDIEPEPQRTISYDIRVLFMDKMGQ	180
DB	642	DITGGNSGSPFDKNGRLGLAFDGNWEAMSGDIEPEPQRTISYDIRVLFMDKMGQ	701
QY	181	CPRLIQELKLI	191
DB	702	CPRLIQELKLI	712

## RESULT 2

US-10-008-355-8

; Sequence 8, Application US/10008355

; Patent No. US20020164759A1

; GENERAL INFORMATION:

; APPLICANT: Travis, James

; APPLICANT: Potempa, Jan S

; APPLICANT: Banbula, Agnieszka

; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101

; CURRENT APPLICATION NUMBER: US/10/008,355

; CURRENT FILING DATE: 2001-11-08

; PRIOR APPLICATION NUMBER: US 60/246,827

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 8

; LENGTH: 699

; TYPE: PRT

; ORGANISM: Porphyromonas gingivalis

US-10-008-355-8

Query Match 93.3%; Score 936; DB 9; Length 699;

Best Local Similarity 100.0%; Pred. No. 1.1e-88;

Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SKSVTAARAQADAMANAAYAEKGRLLFFAGLRMYPGRALPSDANFTMRMSYSGIKGY 60

DB 522 SKSVTAARAQADAMANAAYAEKGRLLFFAGLRMYPGRALPSDANFTMRMSYSGIKGY 581

QY 61 EPQDGAWNYHTTGGVLEKQPKSDEFQAVQENILDLFRTKNYGRYAENGQHLIAFLSN 120

DB 582 EPQDGAWNYHTTGGVLEKQPKSDEFQAVQENILDLFRTKNYGRYAENGQHLIAFLSN 641

QY 121 DTGNGSGSPVDFKNRGLIGLAFDGNWAMSGDIEFEPDLQRTISVDIRYVLFMDKW 178

DB 642 DTGNGSGSPVDFKNRGLIGLAFDGNWAMSGDIEFEPDLQRTISVDIRYVLFMDKW 699

## RESULT 3

US-10-008-355-9

; Sequence 9, Application US/10008355

; Patent No. US20020164759A1

; GENERAL INFORMATION:

; APPLICANT: Travis, James

; APPLICANT: Potempa, Jan S

; APPLICANT: Banbula, Agnieszka

; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101

; CURRENT APPLICATION NUMBER: US/10/008,355

; CURRENT FILING DATE: 2001-11-08

; PRIOR APPLICATION NUMBER: US 60/246,827

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 9

; LENGTH: 720

; TYPE: PRT

; ORGANISM: Porphyromonas gingivalis

US-10-008-355-9

Query Match 42.6%; Score 427.5; DB 9; Length 720;

Best Local Similarity 47.4%; Pred. No. 6.8e-36;

Matches 81; Conservative 34; Mismatches 55; Indels 1; Gaps 1;

QY 22 TEKGRRLFFAGLRMYPGRALPSDANFTMRMSYSGIKGYEPQDGAWNYHTTGGVLEKQ 81

DB 549 ILRAQRTVIAGLLEMDGQDQDFPDANLTLRTFTYGVKGYSPRDNVYVYGHQTTLDGWEKE 608

QY 82 DPKSDEFQAVQENILDLFRTKNYGRYAE-NGQHLIAFLSNNDITGNSGSPVDFKNRGLIG 140

DB 609 DPNWFEVVDPKLKAVERKDFGRYADRSRMPVAFCAATHTTGNNGSGSPVMNANGELIG 668

## QY 141

LAFDGNWAMSGDIEFEPDLQRTISVDIRYVLFMDKWGOCPRLIQELKLI 191

DB 669 LNFDRNWEGVGDDIQYLADYQRSIIIVDIRYVLLVIDKVGCCORLLDEMNIIV 719

## RESULT 4

US-10-008-355-7

; Sequence 7, Application US/10008355

; Patent No. US20020164759A1

; GENERAL INFORMATION:

; APPLICANT: Travis, James

; APPLICANT: Potempa, Jan S

; APPLICANT: Banbula, Agnieszka

; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101

; CURRENT APPLICATION NUMBER: US/10/008,355

; CURRENT FILING DATE: 2001-11-08

; PRIOR APPLICATION NUMBER: US 60/246,827

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 7

; LENGTH: 716

; TYPE: PRT

; ORGANISM: Xylella fastidiosa

US-10-008-355-7

Query Match 33.3%; Score 334; DB 9; Length 716;

Best Local Similarity 40.0%; Pred. No. 3.4e-26;

Matches 74; Conservative 40; Mismatches 53; Indels 18; Gaps 5;

QY 8 ARAIQADAMANAAYAEKGRLLFFAGLRMYPGRALPSDANFTMRMSYSGIKGYEPQDGAW 67

DB 546 ARPIYQLALAD-YNKSHGK-----FVYP-----DANSLRITFGHVKGYSRDKGVE 590

QY 68 YNYHTTGGVLEKQPKSDEFQAVQENILDLFRTKNYGRYAE--GQHLIAFLSNNDITGG 125

DB 591 YTPFTTLQGVMAK-NTGVEPDPSPKSLINAIKAKSVANLADQRICTVVPVNFSLDLDTGG 649

QY 126 NSGSPVDFKNRGLIGLAFDGNWAMSGDIEFEPDLQRTISVDIRYVLFMDKWGOCPRLI 185

DB 650 NSGSPVLDAGHKLGLVGLAFDGNWESVSNWVDFPVMTRTIAVDSRYVQWIMTEVAPAHILL 709

QY 186 QELKL 190

DB 710 KEELN 714

## RESULT 5

US-10-008-355-3

; Sequence 3, Application US/10008355

; Patent No. US20020164759A1

; GENERAL INFORMATION:

; APPLICANT: Travis, James

; APPLICANT: Potempa, Jan S

; APPLICANT: Banbula, Agnieszka

; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101

; CURRENT APPLICATION NUMBER: US/10/008,355

; CURRENT FILING DATE: 2001-11-08

; PRIOR APPLICATION NUMBER: US 60/246,827

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 3

; LENGTH: 52

; TYPE: PRT

; ORGANISM: Porphyromonas gingivalis

US-10-008-355-3

Query Match 27.3%; Score 274; DB 9; Length 52;

Best Local Similarity 100.0%; Pred. No. 1.8e-21;

```

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 123 TGGSGSPVFDKNGRLICLAFDGNWAMSGDIEFEPDLQRTISVDIRVLFEM 174
      |||||
      1 TGGSGSPVFDKNGRLICLAFDGNWAMSGDIEFEPDLQRTISVDIRVLFEM 52
      |||||

RESULT 6
US-10-008-355-6
; Sequence 6, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbulla, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-6

```

```

Query Match      26.2%; Score 262.5; DB 9; Length 732;
Best Local Similarity 30.0%; Pred. No. 9.2e-19;
Matches 67; Conservative 37; Mismatches 86; Indels 33; Gaps 4;

Qy      1 SKSVIAAARAJADAMANAYA-----IEKGKRLFFAGLREMPGRALP--SDANFTMR 51
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      509 SDPFLRAVALYDYNMAQEKAEKILAGKLSSTARPAYMAAVIDYKANNWPVPDANGTLR 568
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      52 MSYSGIKGYEQDGAWYHYHTTGKVLGKDPKSDFAQVENILDLFTKNGRY----- 106
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      569 ISYGMVDGQSDALYKOPFRLDGIVAKH-TGVEPYNAPKKLLDAISVQRFGDHLKSV 627
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      107 -----AENGOLHIAFLSNNDITGGNSGSPVPDKNGRLTCLAFDGNWE 148
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      628 YQDPRGWICRLFSCLDKPFENSVPVNFLSSVDTTGGNSGSPVFNCKELVGLNFDSTIYE 687
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      149 AMSGDIEFEPDLORTISVDIRYLFMIDKWQCPRLIQELKLI 191
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      688 AITDKWFENPPIITRAVHVHVDIYILWMDDEVDAHNLKELDLIV 730
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-10-008-355-5
; Sequence 5, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-5

```

Query Match 25.5%; Score 255.5; DB 9; Length 734;

[illegible]

	Query Match	8.1%	Score 81.5;	DB 10;	Length 240;
	Best Local Similarity	23.0%;	Pred. No. 1.3;		
	Matches: 34;	Conservative 22;	Mismatches 59;	Indels 33;	Gaps 6;
Qy	37	YPGRALPSDANFTMRMSVGSIKGYEPODGAW-YNYHTTGK-----VLEKODPKSDEFAY	90		
Db	108	YPGKEDVSIOVEERATERGPKGFNFNDNVTPFYAKAGAGRIKVIQYPPHYKKNKYVL	167		
Qy	91	QENTILDLFRTKN----YGRYAENGQLIHAFLSNNDITGGNSGSPVFKNGKRLIGLAF---	143		
Db	168	HESTCPVMSVEGSSIVYSAHTES-----GNSCSPVLANSNNELIGIHASD	212		
Qy	144	---DCNWEAMSGDIEFEPDQLQRTTSVDI	168		
Db	213	VKNDNRRNAYG--VYFTPEIKKFAENI	238		



```

; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-008-355-4

Query Match          7.8%; Score 78; DB 9; Length 52;
Best Local Similarity 31.4%; Pred. No. 0.4;
Matches 16; Conservative 13; Mismatches 22; Indels 0; Gaps 0;

QY 123 TGGNSGSPVDFKNGRLGGLAFDGNWAMSGDIEFEPDLQRTISVDIRVLF 173
      ||||| ||||| : : : : : : : : : : : : : : : : : : : :
Db 1 TGGNSGSPVFNKNEVIGIHWGVPNEFGAVFINENRFLKONIEDIHF 51

RESULT 13
US-09-981-286A-33
; Sequence 33, Application US/09981286A
; Publication No. US20020192799A1
; GENERAL INFORMATION:
; APPLICANT: Watowich, Stanley J.
; APPLICANT: Weaver, Scott C.
; APPLICANT: Davey, Robert A.
; TITLE OF INVENTION: Drug Discovery Methods
; FILE REFERENCE: 265.00260101
; CURRENT APPLICATION NUMBER: US/09/981,286A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/240,187
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Adaptein-1 amino acid sequence
US-09-981-286A-33

Query Match          7.7%; Score 77.5; DB 9; Length 165;
Best Local Similarity 24.4%; Pred. No. 2.1;
Matches 42; Conservative 22; Mismatches 63; Indels 45; Gaps 11;

QY 42 LPDANFTMRMSYSGKGYEPQDGA--WYNYHTTGK-----GVLEKQDPKSD-EFA-V 90
      ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 4 LESDKTFPI-MLEGGKINGACVGVGKLFPHVVEGKIDNDVLAALKTKKASKYDLEYADV 62

QY 91 QENI-LDLFR--TKNGRYA-----ENGQHLIAFLSNNDITGNSGSPVDFKNGRL 138
      || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 63 PQNMRADTFKTYTHEKPGCYISWHHGAQVOYENGRTV---PRGVAKGSDGRPILDNOGRV 119

QY 139 IGLAFDG-----NWEAMSGDIEFEPDLQRTISVDIRVLFEMDKW 178
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 120 VAVILGGVNEGSRALSVMWNKLS-----PHYAQLEGVTVKTPENCEQW 165

RESULT 14
US-09-767-041-21
; Sequence 21, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; APPLICANT: Nair, Manoj
; APPLICANT: Chen, Seiyu
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes
; FILE REFERENCE: DEX-0214
; CURRENT APPLICATION NUMBER: US/09/909,567B
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 60/219,834
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-909-567B-44

Query Match          7.6%; Score 76; DB 9; Length 889;
Best Local Similarity 24.1%; Pred. No. 27;
Matches 34; Conservative 18; Mismatches 51; Indels 38; Gaps 8;

QY 1 SKSVIAARAATQADAMANAAYAE-----KGRRLFFAGLREMYPGRALPSDANFTMRM 52
      || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 772 SEETIAL-----VDFEQQAWPLEGFGGALEMKGRRLDLQGIKRVLKG---PDQG--VARS 821

QY 53 S-YGSIKGYEPDQGAWYNYHTTGKGVLEKQDPKSDFAVQENILDLFRTKNYGRVAENQ 111
      ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 822 SCYGDCESEDEATEWITFQ-----VKRVKPKGD-----HKKTPGKRVETGQ 864

QY 112 LHIA--FLSNNDITGNSGSP 130
```

Db 865 TENGHRYOANLEITGPKVASP 885

Search completed: May 16, 2003, 14:03:02  
Job time : 18.3636 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2003, 15:31:51 ; Search time 14.1716 Seconds  
(without alignments)  
(1795.900 Million cell updates/sec)

Title: us-10-008-355-2\_COPY\_522\_712

Perfect score: 1003

Sequence: 1 SKSVIAARAIAQADAMANY.....LFMDKWGCPRLIQELKLI 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1003	100.0	712	23	AA015205
2	274	27.3	52	23	AA015205
3	90	9.0	711	17	AA088649
4	89.5	8.9	756	22	AA095249
5	89.5	8.9	856	22	AA094310
6	89.5	8.9	866	22	AA040262
7	89.5	8.9	923	22	AA042048
8	89.5	8.9	3421	22	AA084902
9	86	8.6	1060	23	AA033805
10	85.5	8.5	278	21	AA040033

11	85.5	8.5	321	21	AA040032	Arabidopsis thalia
12	85.5	8.5	323	21	AA040031	Arabidopsis thalia
13	84	8.4	537	18	AAW22220	Protein encoded by
14	83	8.3	201	22	AA082591	S. epidermidis ope
15	83	8.3	360	23	ABP39512	Staphylococcus epi
16	82	8.2	870	21	AA041742	Human ORFX ORF1506
17	81.5	8.1	240	22	AAU33855	Staphylococcus aur
18	81.5	8.1	254	22	AAU36684	Staphylococcus aur
19	81.5	8.1	309	23	ABP25673	Streptococcus poly
20	81	8.1	283	21	AAV71478	Ehrlichia canis im
21	81	8.1	283	23	AAU96101	Ehrlichia canis p2
22	80.5	8.0	532	17	AA091035	Recombinant V8 pro
23	80.5	8.0	532	18	AAW22219	Protein encoded by
24	80.5	8.0	1317	22	ABG18723	Novel human diagn
25	80.5	8.0	4643	22	AB071609	Drosophila melanog
26	80	8.0	336	13	AA029644	Protease from S. A
27	80	8.0	344	17	AA091033	Beta-galactosidase
28	80	8.0	344	18	AAW22218	Protein encoded by
29	80	8.0	357	13	AA026842	Protease from S. A
30	80	8.0	392	17	AA091034	Beta-galactosidase
31	79.5	7.9	389	23	AB053562	Lactococcus lactis
32	79	7.9	571	18	AAW16399	Konjak mosaic viru
33	78.5	7.8	498	21	AAV75500	Neisseria meningit
34	78	7.8	52	23	AA015207	Staphylococcus aur
35	78	7.8	215	17	AA091044	V8 mature protease
36	78	7.8	509	22	AAW25688	Human protein sequ
37	78	7.8	684	22	AA093433	Human polypeptide,
38	78	7.8	781	21	AA031213	Arabidopsis thalia
39	78	7.8	803	21	AA031212	Arabidopsis thalia
40	78	7.8	846	21	AA031211	Arabidopsis thalia
41	78	7.8	870	22	AA094408	Human protein sequ
42	78	7.8	3432	21	AB07037	Japanese encephali
43	77.5	7.7	971	19	AAW48896	Candida albicans C
44	77	7.7	180	23	AAU75485	S. aureus antigeni
45	77	7.7	332	21	AA068962	Cps2J protein whic

ALIGNMENTS

RESULT 1  
AA015205  
ID AA015205 standard; Protein; 712 AA.  
XX  
AC AA015205;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).  
XX  
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
KW periodontitis.  
XX  
OS Porphyromonas gingivalis.  
XX  
PN WO200238742-A2.  
XX  
PD 16-MAY-2002.  
XX  
PF 08-NOV-2001; 2001WO-US46782.  
XX  
PR 08-NOV-2000; 2000US-246827P.  
XX  
PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
PI Travis J, Potempa JS, Banbula A, Bugno M;  
XX  
DR WPI: 2002-490075/52.  
XX  
PT N-PSDB: AAL43635.  
XX  
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
the dipeptidylpeptidase for protecting an animal from periodontal

```
PT disease caused by Porphyromonas gingivalis
XX Claim 7; Fig 4; 65pp; English.
XX
XX The invention comprises the amino acid and coding sequence of the
XX Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
XX enzyme has amidolytic activity for cleavage of a peptide bond between the
XX second and third amino acids from the N-terminal end of a target peptide.
XX The DPP-7 target peptide has an aliphatic or aromatic residue as a
XX substituent on the alpha-carbon atom of the second amino acid from the
XX N-terminal end. The DPP-1 protein and DNA sequences of the invention are
XX useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
XX reducing the growth of a bacterium and protecting an animal from a
XX periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
XX or periodontitis). The present amino acid sequence represents the
XX Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the
XX invention.
XX
XX Sequence 712 AA;
XX
XX Query Match 100.0%; Score 1003; DB 23; Length 712;
XX Best Local Similarity 100.0%; Pred. No. 9 6e-101;
XX Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 SKSVIAARAATQADAMANAYAEKGRLEFFAGLRMPYGRALPSDANFTMRMSYGIKGY 60
XX Db 522 SKSVIAARAATQADAMANAYAEKGRLEFFAGLRMPYGRALPSDANFTMRMSYGIKGY 581
XX
XX Qy 61 EPDQGANVNYHTTGGVLEKQDKPSDEFVQENILDLFRTKNYGRYAENGQHLIAFLSN 120
XX Db 582 EPDQGANVNYHTTGGVLEKQDKPSDEFVQENILDLFRTKNYGRYAENGQHLIAFLSN 641
XX
XX Qy 121 DITGNSGSPVFDKNGRLIGLAFDGNWAMSGDIEFEFDLQRTISVDIRYVLFMDKKGQ 180
XX Db 642 DITGNSGSPVFDKNGRLIGLAFDGNWAMSGDIEFEFDLQRTISVDIRYVLFMDKKGQ 701
XX
XX Qy 181 CPRLIQELKLI 191
XX Db 702 CPRLIQELKLI 712
XX
XX RESULT 2
XX AA015206
XX ID AA015206 standard; Protein: 52 AA.
XX AC AA015206;
XX
XX DT 05-SEP-2002 (first entry)
XX
XX DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.
XX
XX KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
XX KW DPP-7 inhibitor identification; periodontal disease; gingivitis;
XX KW periodontitis.
XX
XX OS Porphyromonas gingivalis.
XX
XX PN WO200238742-A2.
XX
XX PD 16-MAY-2002.
XX
XX XX 08-NOV-2001; 2001WO-US46782.
XX
XX XX 08-NOV-2000; 2000US-246827P.
XX
XX XX (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX PI Travis J, Potempa JS, Banbula A, Bugno M;
XX XX WPI; 2002-490075/52.
XX
XX PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
XX the dipeptidylpeptidase for protecting an animal from periodontal
```

```
PT disease caused by Porphyromonas gingivalis
XX Example 6; Fig 5; 65pp; English.
XX
XX The invention comprises the amino acid and coding sequence of the
XX Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
XX enzyme has amidolytic activity for cleavage of a peptide bond between the
XX second and third amino acids from the N-terminal end of a target peptide.
XX The DPP-7 target peptide has an aliphatic or aromatic residue as a
XX substituent on the alpha-carbon atom of the second amino acid from the
XX N-terminal end. The DPP-1 protein and DNA sequences of the invention are
XX useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
XX reducing the growth of a bacterium and protecting an animal from a
XX periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
XX or periodontitis). The present amino acid sequence represents the
XX Porphyromonas gingivalis dipeptidylpeptidase-7
XX (DPP-7) enzyme.
XX
XX Sequence 52 AA;
XX
XX Query Match 27.3%; Score 274; DB 23; Length 52;
XX Best Local Similarity 100.0%; Pred. No. 1 7e-22;
XX Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 123 TGGNSGSPVFDKNGRLIGLAFDGNWAMSGDIEFEFDLQRTISVDIRYVLFM 174
XX Db 1 TGGNSGSPVFDKNGRLIGLAFDGNWAMSGDIEFEFDLQRTISVDIRYVLFM 52
XX
XX RESULT 3
XX AAR88649
XX ID AAR88649 standard; Protein: 711 AA.
XX AC AAR88649;
XX
XX DT 02-SEP-1996 (first entry)
XX
XX DE Neisseria meningitidis B2163 transferrin receptor Tbp2 subunit.
XX
XX KW Transferrin receptor; Tbp2 subunit; deletion mutant; vaccine;
XX KW passive immunisation; immunotherapy; IM2169; IM2394.
XX
XX OS Neisseria meningitidis (strain B2163).
XX
XX FH Key Location/Qualifiers
XX FT Protein 1..711
XX FT Peptide /label= Tbp2_subunit
XX FT 1..20 /label= signal_peptide
XX FT Domain 21..366
XX FT /label= Domain_1
XX FT /note= "N-terminal domain"
XX FT Domain 367..564
XX FT /label= Domain_2
XX FT /note= "hinge domain"
XX FT Domain 565..711
XX FT /label= Domain_3
XX FT /note= "C-terminal domain"
XX
XX PN WO9533049-A2.
XX
XX PD 07-DEC-1995.
XX
XX PF 30-MAY-1995; 95WO-FR00701.
XX
XX PR 31-MAY-1994; 94FR-0006594.
XX
XX PA (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
XX PA (TRGE ) TRANSGENE SA.
XX
XX PI Jacobs E, Legrain M, Mazarin V, Lissolo L;
XX PI Millet MB;
XX
```

DR WPI: 1996-030562/03.  
XX DR N-PSDB; AAT11244.  
XX PT Polypeptide(s) for vaccination against *Neisseria meningitidis* group  
PT B - comprising deletion mutants of transferrin receptor Tbp2  
PT subunit  
XX  
PS Disclosure: Page 82-87; 114pp; French.  
XX  
CC The present sequence is that of the *N.meningitidis* strain B2163  
CC transferrin Tbp2 subunit. The Tbp2 polypeptide has three  
CC domains (see features table); deletion mutants in which at least  
CC one of the domains is partially or totally deleted are claimed,  
CC provided that the first and second domains are not simultaneously  
CC partially or totally deleted. The positions of the 3 domains in  
CC B2163 are defined by alignment with the IM2169 sequence. The  
CC deletion mutant polypeptides of the invention can generate an  
CC immune response against *N.meningitidis*.  
XX  
SQ Sequence 711 AA:  
Query Match 9.0%; Score 90; DB 17; Length 711;  
Best Local Similarity 21.8%; Pred. No. 1.2;  
Matches 38; Conservative 28; Mismatches 56; Indels 42; Gaps 7;  
QY 2 KSVIAARATQADAMANAYAIKGRKLFAGLRMPGRLPSDANFTMRMSGYSGIKGYE 61  
Db 518 KNSKSAQAGESSQADAKTEQVQSQMFLQ--ERTDEKIPSEQNIVR----- 565  
QY 62 PQGAWNY---HTTGGK-VLEKQDPKSDFAVOENILDLFRKNYGRYAENGQHLIAFL 117  
Db 566 ---GSWYGHIASTSGNSGASDREGGNRAEFTV-----NFGKKITGL----- 606  
QY 118 SNNDITGNGSGSPVFDKNGRLIGLAFDGNWEAMSGDIEFEPDLQRTISVDIRV 171  
Db 607 -----TAENRQEAFTFDGKIEGNGSGT--AKTAEGLFDLQKNTTRTPKAVI 653  
RESULT 4  
AAB95249  
ID AAB95249 standard; Protein; 756 AA.  
XX AAB95249;  
XX  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:17413.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI: 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 17413; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 756 AA;  
Query Match 8.9%; Score 89.5; DB 22; Length 756;  
Best Local Similarity 25.1%; Pred. No. 1.5;  
Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;  
QY 19 AYALEKGRKLFAGLRMPGRLPSDA--NFTMRM-----SYGSIKGYEPDQGAWNYH 71  
Db 149 SYGVRGRVCFEMKINEEISVKHLPLSTPPDPHVVRIGWSLDSCSTQLGEEFFS---YGYG 205  
QY 72 TTGKGVLEKQDPKSDFAVOENILDLFRKNYGRYAENGQHLIAFLSNNDITG----- 124  
Db 206 GTCK-----KSTN-----SRFENYGDKFAE-----NDVIGCFADFE 236  
QY 125 -GNSGSPVFDKNGRLIGLAFDGNWEAMSGDIEFEPDLQRTISVDIRV 170  
Db 237 CGNDVELSFTKNGKWMGIAFRIOKEALGGOALYPHVLVKNCAVEFNF 283  
RESULT 5  
AAB94310  
ID AAB94310 standard; Protein; 856 AA.  
XX AAB94310;  
XX  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:14779.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300253.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX







```
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.5%; Score 85.5; DB 21; Length 278;
Best Local Similarity 27.6%; Pred. No. 0.93;
Matches 42; Conservative 16; Mismatches 67; Indels 27; Gaps 8;

Qy 45 DANETMRMSYGSIKGYEP-QDGAWNYHTGK-----GVLEKODPKSDEPAV-----Q 91
Db 123 DAKGTRFSKECKIVGLDPDNDLAVLKIEGRELNPVVLGTSNDLRFVQSCFAIGNPYGY 182
Qy 92 ENILDLFRTKNYGRY--AENGQ-LHIAFLSNNDITGNSGSPVFKNGRLIGL---AFDQ 145
Db 183 ENTLTIGVVSGLGREIPSPNGKSISEAQTODADINSGSGPLDLSYGHTICVNTATFTR 242
Qy 146 NWEAMSGDIEFEPDLQRTISVD--IRYVLFMI 175
```

Db 243 KSGMSSGVNF-----AIPIDTVVRTVPYLI 268  
RESULT 11  
AAG40032  
ID AAG40032 standard; Protein; 321 AA.  
XX AAG40032;  
XX  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 49615.  
XX  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 05-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.



```
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.5%; Score 85.5; DB 21; Length 321;
Best Local Similarity 27.6%; Pred. NO. 1.2;
Matches 42; Conservative 16; Mismatches 67; Indels 27; Gaps 8;

Qy 45 DANFTMRSYGSIKGYEP-QDGAWNYHTGK-----GVLEKQDPKSDFAV-----Q 91
Db 166 DAKGTRFKEKGIVGLDNDLAVLKIETEGRELPVVLGTSNDLRVGCSCFAIGNPYGY 225
Qy 92 ENILDFRTKNGRY--AENGQ-LHIAFLSNDITGGNSGSPVDFKNGRLIGL---AFDG 145
Db 226 ENTLLGVVSGLRIPSPNGKSGISEATQTDADINSNGSGGLDLSYGHITGVNTATFTTR 285
Qy 146 NWEAMSGDIEFDPQLQRTISVD--TRYVLFI 175
```

```
Db 286 KSGMSGCVNF-----AIPDITVVRTVPYLI 311
RESULT 12
AAG40031
ID AAG40031 standard; Protein: 323 AA.
XX
AC AAG40031;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49614.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
```

PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	24-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-AUG-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 8.5%; Score 85.5; DB 21; Length 323;		
Best Local Similarity 27.6%; Pred. No. 1.2;		
Matches 42; Conservative 16; Mismatches 67; Indels 27; Gaps 8;		
Qy	45	DANFTMRMSYSGIKGYEP-QDCAWNYHTTK-----GVLEKQDPKSDFAV-----Q 91
Db	168	DAKGRFSKEGRVGLDPDNDLAVLKUETEGRELNPVVLGTSDNLRVQSCFAIGNPYGY 227
Qy	92	ENILDLFRTKNYGRY--AENGQ-LHIAFLSNNDITGNSGSPVFDKNGRLIGL----AFDG 145
Db	228	ENTLTIGVVSGLGREIPSPNGKKSISEAIQTADINSNGSGGLLDSYGHITGVNTATFTR 287

Query Match	8.4%;	Score 84;	DB 18;	Length 537;
Best Local Similarity	28.1%;	Pred. No. 3.6;		
Matches	16;	Conservative	16;	Mismatches 24; Indels 6; Gaps 1;
QY	123	TGGNSGSPVDFKNGRLICLAFDGNWEAMSGDIEPEPLORTISVDIRVLEMI-----D	176	
Db	289	TGGNSGSPVDFKNGRLICLAFDGNWEAMSGDIEPEPLORTISVDIRVLEMI-----D	176	
QY	177	KWGO 180		
Db	349	RWGR 352		
RESULT 14				
AAG82591	ID	AAG82591 standard; Protein; 201 AA.		
XX	AC	AAG82591;		
XX	DT	03-SEP-2001 (first entry)		
XX	DE	S. epidermidis open reading frame protein sequence SEQ ID NO:2276.		
XX	KW	Staphylococcus epidermidis SRI strain; infection; diagnosis;		
XX	KW	vaccination; endocarditis.		
XX	OS	Staphylococcus epidermidis.		
XX	OS	WO200134809-A2.		
XX	PN	17-MAY-2001.		
XX	PD	09-NOV-2000; 2000WO-US30782.		
XX	PF	09-NOV-1999; 99US-0164258.		
XX	PR	(GLAX ) GLAXO GROUP LTD.		
XX	PA	Kimmerly WJ;		
XX	PI	WPI; 2001-316495/33.		
XX	DR	N-PSDB; AAH53441.		
XX	DR			
XX	PT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,		
XX	PT	useful for vaccinating against infections, e.g. endocarditis -		
XX	PS	Claim 18; Page 606; 2188pp; English.		
XX	CC	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides		
XX	CC	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.		
XX	CC	(I) and (II) can have antibacterial activity and therefore can be used		
XX	CC	in vaccination. The nucleic acids (I) may be used to produce the		
XX	CC	S. epidermidis polypeptides (II) via the production of vectors		
XX	CC	containing them which are used to produce hosts cells which express the		
XX	CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be		
XX	CC	used to vaccinate subjects and to raise antibodies against the bacteria.		
XX	CC	The polypeptides may also be used to assay for other inhibitors of their		
XX	CC	activity and therefore identify compounds that may be used for the		
XX	CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to		
XX	CC	AAH5090 represent specifically claimed S. epidermidis genomic DNA		
XX	CC	polynucleotide sequences from the present invention. AAH55091 to		
XX	CC	AAH55098 represent oligonucleotide sequences and primers which are used		
XX	CC	in the exemplification of the present invention.		
XX	CC	N.B. The present invention specifically claims all the polynucleotide		
XX	CC	sequences given in the sequence listing of the present specification,		
XX	CC	however the sequence listing only goes up to SEQ ID NO:4454 so even		
XX	CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,		
XX	CC	no sequences are present for SEQ ID NO:4455 to 4464.		
XX	SQ	Sequence 201 AA;		



Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	84	8.4	537	1	US-08-657-192-15	Sequence 15, Appl
2	83	8.3	360	4	US-09-134-001C-4357	Sequence 4357, Appl
3	82	8.2	872	1	US-08-491-357-3	Sequence 3, Appl
4	82	8.2	872	3	US-08-968-633-3	Sequence 3, Appl
5	82	8.2	872	3	US-09-196-466-3	Sequence 3, Appl
6	82	8.2	872	5	PCF-US96-10823-3	Sequence 3, Appl
7	81	8.1	283	4	US-09-660-587-4	Sequence 4, Appl
8	81	8.1	283	4	US-09-261-358A-4	Sequence 4, Appl
9	80.5	8.0	532	1	US-08-657-192-9	Sequence 9, Appl
10	80.5	8.0	532	3	US-08-523-373-7	Sequence 7, Appl
11	80	8.0	344	1	US-08-657-192-3	Sequence 3, Appl
12	80	8.0	344	3	US-08-523-373-5	Sequence 3, Appl
13	80	8.0	392	3	US-08-523-373-6	Sequence 6, Appl
14	78	7.8	215	3	US-08-523-373-24	Sequence 24, Appl
15	75	7.5	213	3	US-08-523-373-22	Sequence 22, Appl
16	75	7.5	214	3	US-08-523-373-23	Sequence 23, Appl
17	74	7.4	793	3	US-08-433-522A-10	Sequence 10, Appl
18	74	7.4	793	3	US-09-135-166-10	Sequence 10, Appl
19	74	7.4	793	4	US-08-942-046-10	Sequence 10, Appl
20	73.5	7.3	977	4	US-09-206-942-53	Sequence 53, Appl
21	73.5	7.3	983	4	US-09-206-942-51	Sequence 51, Appl
22	72.5	7.2	280	3	US-08-733-230-4	Sequence 4, Appl
23	72.5	7.2	280	4	US-08-953-326-4	Sequence 4, Appl
24	72.5	7.2	663	1	US-08-765-081-7	Sequence 7, Appl
25	72.5	7.2	663	3	US-09-098-082-7	Sequence 7, Appl
26	72.5	7.2	691	1	US-08-064-174-2	Sequence 2, Appl
27	72.5	7.2	691	1	US-08-066-167-4	Sequence 4, Appl

```
Db    289 TGGNSGSPVFNEKKEVIGHHGGVPNFGAVFINENVRNLQKNLIEDIHFELRLYRRHH   348  
      :|||:|||::: ||::: | :|| :::: |:| :| :| :| :|  
  
Qy    177 KWQG 180  
      :||:  
  
Db    349 RWGR 352  
      :||:  
  
RESULT 2  
US-09-134-001C-4357  
; Sequence 4357, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: CPC-007  
; CURRENT APPLICATION NUMBER: US/09/134_001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4357  
; LENGTH: 360  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4357
```

Query Match                  8.3%; Score 83; DB 4; Length 360;  
Best Local Similarity      24.5%; Pred. No. 0.66;  
Matches    34; Conservative    15; Mismatches    46; Indels     44; Gaps       6;

```
Qy     57 IKGYEPDGAWNYHTTCKGVLEKQDPKSDEFVAOVENILDLFRT-KNY-----CRYAEN 109  
       :|| |          ||::| |::| |::| |::| |::| |::| |::| |::| |::| |  
Db     20 MKGAHPMKNIKKPDLKGSKLLKEYDLTGEEF---EGLIDFAMTLKRYKQQGTPHYLEG 76  
       :|| |          ||::| |::| |::| |::| |::| |::| |::| |::| |  
Qy     110 GOL-----HIATLSNNDTITGGNSGSPVKNGRILIGLAFDG 145  
        :|           ||| ||| ||| ||| ||| ||| ||| |::| |::| |  
Db     77 KNIALLFKTSRTRAFTVASIDLGAHEFLGNKDIOLGKES--VEDTAKVLGRMFDG 134  
        :|           ||| ||| ||| ||| ||| ||| ||| ||| |::| |  
Qy     146 NWEAMSGDIIEFPDLORTI 164  
         ||| |||  
Db     135 -----IEPRGFSEQTV 145  
         ||| |||  
  
RESULT 3  
US-08-491-357-3  
; Sequence 3, Application US/08491357  
; Patent No. 5716782  
; GENERAL INFORMATION:  
; APPLICANT: Golemis, Erica A.  
; APPLICANT: Law, Susan  
; APPLICANT: Estojak, JoAnne  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL  
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL  
; TITLE OF INVENTION: ALTERATIONS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
; STREET: 1601 Market Street Suite 720  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/491.357
```

```
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. 6100384 Relevant
; TOPOLOGY: No. 6100384 Relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-968-633-3

Query Match      8.2%; Score 82; DB 3; Length 872;
Best Local Similarity 24.3%; Pred. No. 3.1;
Matches 43; Conservative 18; Mismatches 60; Indels 56; Gaps 9;

QY 3 SVIAARAIOADAMANAIAIEKGRLEFAGLREMPG-----RAL 42
Db 580 TLVACSRAPVEDAKQLASFLHGNASLLFRRTKAPCGPGEGLSSLLHNPDKASSIQSRPL 639

QY 43 PSDANFTMRMSYSGSIKG-YEPQDGAW---YNY-HTTGKGVLEKQDPKSDFAVOENILDL 97
Db 640 PPKPKFT---SODSPDGOYENSEGGWMEDYDVHLQKEEFK-----TKELLEK 687

QY 98 FRTKNYGRYAENGOLHTAFLS-----NNDITGNSGSPVPDKNGRLIGL 141
Db 688 GNIVROGK---GQLEQLQKQFERLEQEVSRPIDHDLANNTPAQPLVP--GRTGGL 738

RESULT 5
US-09-196-466-3
; Sequence 3, Application US/09196466
; Patent No. 6124434
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,466
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-968-633-3

Query Match      8.2%; Score 82; DB 3; Length 872;
Best Local Similarity 24.3%; Pred. No. 3.1;
Matches 43; Conservative 18; Mismatches 60; Indels 56; Gaps 9;

QY 3 SVIAARAIOADAMANAIAIEKGRLEFAGLREMPG-----RAL 42
Db 580 TLVACSRAPVEDAKQLASFLHGNASLLFRRTKAPCGPGEGLSSLLHNPDKASSIQSRPL 639

QY 43 PSDANFTMRMSYSGSIKG-YEPQDGAW---YNY-HTTGKGVLEKQDPKSDFAVOENILDL 97
Db 640 PPKPKFT---SODSPDGOYENSEGGWMEDYDVHLQKEEFK-----TKELLEK 687

QY 98 FRTKNYGRYAENGOLHTAFLS-----NNDITGNSGSPVPDKNGRLIGL 141
Db 688 GNIVROGK---GQLEQLQKQFERLEQEVSRPIDHDLANNTPAQPLVP--GRTGGL 738

RESULT 6
PCT-US96-10823-3
; Sequence 3, Application PC/TUS9610823
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10823
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US96-10823-3

Query Match      8.2%; Score 82; DB 5; Length 872;
Best Local Similarity 24.3%; Pred. No. 3.1;
Matches 43; Conservative 18; Mismatches 60; Indels 56; Gaps 9;

QY 3 SVIAARAIOADAMANAIAIEKGRLEFAGLREMPG-----RAL 42
Db 580 TLVACSRAPVEDAKQLASFLHGNASLLFRRTKAPCGPGEGLSSLLHNPDKASSIQSRPL 639

QY 43 PSDANFTMRMSYSGSIKG-YEPQDGAW---YNY-HTTGKGVLEKQDPKSDFAVOENILDL 97
```

Db 640 PSPPKFT---SQDSPDGOYENSEGWMEDYDVHLOGKEEFK-----TQKELLEK 687  
Qy 98 FRKNYGRYAENGOLHIAFLS-----NNDITGNSGSPVFDKNGRLIGL 141  
Db 688 GNIVROGK-----GQLEQOLKQERLEOEVSRRPIDHDLANWTAPQPLVP--GRTGGL 738  
RESULT 7  
US-09-660-587-4  
; Sequence 4, Application US/09660587  
; Patent No. 6392023  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP2  
; CURRENT APPLICATION NUMBER: US/09/660.587  
; CURRENT FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: 09/261.358  
; PRIOR FILING DATE: 1999-03-03  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 4  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of p28-5 protein  
US-09-660-587-4

Query Match 8.1%; Score 81; DB 4; Length 283;  
Best Local Similarity 26.9%; Pred. No. 0.77;  
Matches 43; Conservative 20; Mismatches 67; Indels 30; Gaps 7;  
Qy 23 EKGRKLFAGUREMYPGRALPS---DANFTMR---MSYGSIK--GYEPQDGAWNYHTTG 74  
Db 58 EKKKTVVYVGLKENWAGDAISSQSPDDNFTIRNYSFKYASKNFLGFAVAIG-----YSIG 112  
Qy 75 KGVLEKQDPKSDFAVQENILDLFRTKNYGRYAENGOLHIAFLSNDDITGNSGSP----- 130  
Db 113 SPRIE-----VEMSYEAFDVKNPGDNYKNGAYRYCALSHQDDADDMDTSATDKF 161  
Qy 131 VFDKNGRLIGLAFDGN--WEAMSGDIEPEPDLQRTISVDI 168  
Db 162 VYLINEGLLNISFTWICETASKNIPLSPYICAGIGTDL 201

RESULT 8  
US-09-261-358A-4  
; Sequence 4, Application US/09261358A  
; Patent No. 6403780  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: McBride, Jere W.  
; APPLICANT: Yu, Xue-Jie  
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein  
; FILE REFERENCE: D6152CIP  
; CURRENT APPLICATION NUMBER: US/09/261.358A  
; CURRENT FILING DATE: 1999-03-03  
; PRIOR APPLICATION NUMBER: 09/201.458  
; PRIOR FILING DATE: 1998-11-30  
; NUMBER OF SEQ ID NOS: 33  
; SEQ ID NO 4  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: Ehrlichia canis  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence of Eca28SA2 protein  
US-09-261-358A-4

Query Match 8.1%; Score 81; DB 4; Length 283;  
Best Local Similarity 26.9%; Pred. No. 0.77;  
Matches 43; Conservative 20; Mismatches 67; Indels 30; Gaps 7;  
Qy 23 EKGRKLFAGUREMYPGRALPS---DANFTMR---MSYGSIK--GYEPQDGAWNYHTTG 74  
Db 58 EKKKTVVYVGLKENWAGDAISSQSPDDNFTIRNYSFKYASKNFLGFAVAIG-----YSIG 112  
Qy 75 KGVLEKQDPKSDFAVQENILDLFRTKNYGRYAENGOLHIAFLSNDDITGNSGSP----- 130  
Db 113 SPRIE-----VEMSYEAFDVKNPGDNYKNGAYRYCALSHQDDADDMDTSATDKF 161  
Qy 131 VFDKNGRLIGLAFDGN--WEAMSGDIEPEPDLQRTISVDI 168  
Db 162 VYLINEGLLNISFTWICETASKNIPLSPYICAGIGTDL 201

RESULT 9  
US-08-657-192-9  
; Sequence 9, Application US/08657192  
; Patent No. 5747321  
; GENERAL INFORMATION:  
; APPLICANT: YABUTA, Masayuki  
; APPLICANT: OHSUYE, Kazuhiro  
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8  
; TITLE OF INVENTION: PROTEASES  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/657.192  
; FILING DATE: 03-JUN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-170086  
; FILING DATE: 02-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 001560-264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 532 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-657-192-9

Query Match 8.0%; Score 80.5; DB 1; Length 532;  
Best Local Similarity 30.5%; Pred. No. 2.2;  
Matches 18; Conservative 15; Mismatches 25; Indels 1; Gaps 1;  
Qy 123 TGGNSGSPVFDKNGRLIGLAFDGNWAMSGDIEPEPDLQRTISVDIRYVLF-MIDKMQQ 180  
Db 289 TGGNSGSPVFNKNEVIGIHWGGVFNENGVAFNENVRNFKQNIEDRLYRRHRWGR 347

RESULT 10  
US-08-523-373-7



TELEFAX: 703-836-2021

COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk

MEDIUM TYPE: FLOPPY disk

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373--5

Query Match      8.0%; Score 80; DB 3; Length 344;
Best Local Similarity 31.5%; Pred. No. 1.3;
Matches 17; Conservative 13; Mismatches 24; Indels 0; Gaps 0;

QY   123 TGGNSGSPVFDKNGRLIGLAFDGNWEAMSGDIEPEPDQLQRTISVDIRYVLFMID 176
      |||||::: :||: | :|: : : : :| :|
Db    289 TGGNSGSPVFNEKKEVGIIHGGVPNEFNGAVFINENVRNFLKNQIEDIHFAND 342

RESULT 13
US-08-523-373-6
; Sequence 6, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-6

Query Match      8.0%; Score 80; DB 3; Length 392;
Best Local Similarity 31.5%; Pred. No. 1.6;
Matches 17; Conservative 13; Mismatches 24; Indels 0; Gaps 0;

QY   123 TGGNSGSPVFDKNGRLIGLAFDGNWEAMSGDIEPEPDQLQRTISVDIRYVLFMID 176
      |||||::: :||: | :|: : : : :| :|
Db    289 TGGNSGSPVFNEKKEVGIIHGGVPNEFNGAVFINENVRNFLKNQIEDIHFAND 342
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: . amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-523-373-22

Query Match 7.5%; Score 75; DB 3; Length 213;
Best Local Similarity 32.6%; Pred. No. 2.3;
Matches 15; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 123 TGGNSGSPVFDKNGRLIGLAFDGNWEAMSGDIEFEPDLQRTISVDI 168
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 165 TGGNSGSPVFNEXNEVIGIHGGVPNEFNGAVFINENVRNFKONI 210

Search completed: May 16, 2003, 13:49:03
Job time : 12.2508 secs
```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2003, 15:31:51 ; Search time 12.0565 Seconds  
(without alignments)  
1522.972 Million cell updates/sec

Title: US-10-008-355-2\_COPY\_522\_712

Perfect score: 1003

Sequence: 1 SKSVIAARAIQADAMANAY.....LFMDKMGQCPRLIQELKLI 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	33.3	716	2 G82627	hypothetical prote
2	95	9.5	507	2 T34152	hypothetical prote
3	93	9.3	513	2 D96980	sensory transducti
4	90	9.0	711	2 S70660	transferrin-bindin
5	89.5	8.9	856	2 T13159	genome polyprotein
6	89.5	8.9	3411	1 GNWVYP	genome polyprotein
7	89.5	8.9	3411	1 GNWVYP	genome polyprotein
8	86.5	8.6	576	2 B39228	flagellin B - Camp
9	86	8.6	1060	2 H86686	DNA-directed DNA p
10	84	8.4	729	2 AF3591	alkaline phosphata
11	84	8.4	968	2 S46992	protein p130 - rat
12	83	8.3	698	2 D81832	transferrin-bindin
13	82.5	8.2	759	2 AE1686	pyruvate formate-l
14	82	8.2	402	2 C87291	hypothetical prote
15	81.5	8.1	493	2 G90604	hypothetical prote
16	81.5	8.1	603	2 T04733	auxin-regulated pr
17	81	8.1	239	2 B89967	serine proteinase
18	81	8.1	663	2 AG0782	colicin I receptor
19	81	8.1	1104	2 A60999	alpha-amylase (EC
20	80.5	8.0	325	2 E75404	ABC transporter, A
21	80.5	8.0	434	2 A29525	methyl coenzyme M
22	80	8.0	336	1 PRSASK	glutamyl endopepti
23	80	8.0	342	2 G98973	hypothetical prote
24	80	8.0	357	2 S21758	glutamic acid-spec
25	80	8.0	2043	2 T18524	scavenger receptor
26	80	8.0	3433	1 GNWVKV	genome polyprotein
27	79.5	7.9	240	2 C89967	serine proteinase
28	79.5	7.9	389	2 H86656	hypothetical prote
29	79.5	7.9	747	2 H82943	hypothetical prote

30	78.5	7.8	289	2 A89824	conserved hypothet
31	78.5	7.8	498	2 H81782	adhesin MafB2 NMA2
32	78.5	7.8	608	2 T34994	probable long-chain
33	78.5	7.8	3388	1 GNWVDP	genome polyprotein
34	78	7.8	271	2 G83780	hypothetical prote
35	78	7.8	846	2 T04533	hypothetical prote
36	78	7.8	1084	2 C82931	hypothetical prote
37	78	7.8	3432	1 GNWVJE	genome polyprotein
38	77.5	7.7	515	2 A70905	hypothetical prote
39	77.5	7.7	647	1 G86118	2',3'-cyclic-nucle
40	77.5	7.7	647	1 G91277	2',3'-cyclic-nucle
41	77.5	7.7	3391	1 GNWV26	genome polyprotein
42	77	7.7	161	2 H84458	hypothetical prote
43	77	7.7	170	2 AD1934	hypothetical prote
44	77	7.7	332	2 C95077	pneumococcal surfa
45	77	7.7	412	2 H75484	hypothetical prote

ALIGNMENTS

RESULT 1

G82627

hypothetical protein XFI887 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: G82627

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: G82627

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-716 <SIM>

A:Cross-references: GB:AE004008; GB:AE003849; MID:g9106961; PIDN:AAF84693.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, U.E.A.; Carraro, D.M.; Carrer

as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramochi, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XFI887

Query Match 33.3%; Score 334; DB 2; Length 716;

Best Local Similarity 40.0%; Pred. No. 2.1e-22;

Matches 74; Conservative 40; Mismatches 53; Indels 18; Gaps 5;

QY 8 ARATQADAMANAYAIKRGKRLFFAGLREMPGRALPDSANFTMRMSYGSIKGYEPQDCAW 67

Db 546 ARPIYQALAD-YNKSHGK-----FVYP-----DANSSLRITFCHVKGYSPKDGVE 590

QY 68 YNYHTTCKGVLEKDKPSDFEFAVENILDLFRKNYGRYAEN--GQLHIAFLSNNDITGG 125

Db 591 YTPPTTLQGVMAK-NTGVPEPDSKLINAIAKSAKSAANLADQRIQTGFVFNFLSLDITGG 649

QY 126 NGSFVFDKNGRLIGLAFDGNWAEAMSGDIEPEPLQRTISVDIRYLFMDIKWQCQCPRLI 185

Db 650 NGSFVLDAHGKLVGLAFDGNWESVSSNWVDFPVTMTTIAVDSRYVGVWITEVAPAPHL 709

QY 186 QELKL 190

||| |









A;Cross-references: GB:AE005673; NID:g13421489; PIDN:AAK22327.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC0340  
C;Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology.

Query Match 8.2%; Score 82; DB 2; Length 402;  
Best Local Similarity 25.0%; Pred. No. 9.6;  
Matches 48; Conservative 27; Mismatches 71; Indels 46; Gaps 10;

QY 3 SVIAAARAICADAMANAAYAJEKGKRLLF-----AGLREMYPGRALPSDANFTMRMSYGSI 57  
| : | | | : : : | | | | : | | | : | | |  
Db 208 SAVNALRAQKYDFEKQHGVKLGFMSFFVXAVVAAL-----KAIP-DVN-----AEI 253

QY 58 KGYPEQPDAWNYHTTG-----KGVLEKDPKSDEF---AVQENILDLFRKNRYGAE 108  
| | | : | : | | | : : : | | : : | | : : | |  
Db 254 DG---QDVIIKHYDIVGAVGTKDKGLVPVRDADALNLAGEIKTIGDL-----GKKRAR 304

QY 109 NGQLHIA-----FLNNITGGNSGPVD--KNGRLLGLAFDGHWAMSGDIEFEPPD 159  
|| || : || | | | : : : | | | : : | | |  
Db 305 NGQLAIEDMOGGTTINTNGGIYSLSMTPLINAPQSGLGMHAIKERPVMVINGKIEIRPM 364

QY 160 LORTISVDIRVV 171  
: | | | | |  
Db 365 MYLAISYDHRIIV 376

RESULT 15  
G90604  
hypothetical protein MYPU\_7430 [Imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C;Species: Mycoplasma pulmonis  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C;Accession: G90604  
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallisson, F.; Moszer,  
Nucleic Acids Res. 29, 2145-2153, 2001  
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma  
A;Reference number: A99512; MUID:21267165; PMID:11353084  
A;Accession: G90604  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-493 <KUR>  
A;Cross-references: GB:AL445566; PID:g14090158; PIDN:CAC13916.1; GSPDB:GN00153  
A;Experimental source: strain UAB CTIP  
C;Genetics:  
A;Gene: MYPU\_7430  
A;Genetic code: GSC3

Query Match 8.1%; Score 81.5; DB 2; Length 493;  
Best Local Similarity 28.6%; Pred. No. 14;  
Matches 16; Conservative 17; Mismatches 20; Indels 3; Gaps 2;

QY 105 RYAENGOLHIAFLSNNDITGCNSGPSVFDRKNRLIGL-AFDGNWEAMSGDIEFEPPD 159  
| : | | | : | : ||| ||| : : : : | : : | | : : |  
Db 378 KYIYSGLLY--YFRNTALSOGSSGSKTVDXNNKIIVGHIALGSDLSGLTGVSFAFKSE 431

Search completed: May 25, 2003, 15:34:50  
Job time : 17.0565 secs

Search completed: May 25, 2003, 15:34:50  
Job time : 17.0565 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2003, 15:31:51 ; Search time 14.1716 Seconds  
(without alignments)  
2777.023 Million cell updates/sec

Title: US-10-008-355-2\_COPY\_522\_712  
Perfect score: 1003  
Sequence: 1 SKSVIAARAIAQADANAY.....LFMDKWGCPRLIQELKLI 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	334	33.3	716	Q9PC94	Q9PC94 xylella fas
2	101.5	10.1	305	Q8XUP6	Q8XUP6 ralstonia s
3	93	9.3	513	Q97LA8	Q97LA8 clostridium
4	90.5	9.0	408	Q96604	Q96604 avian infec
5	90	9.0	711	Q51284	Q51284 neisseria m
6	89.5	8.9	856	Q76022	Q76022 homo sapien
7	89.5	8.9	856	Q9BUJ2	Q9BUJ2 homo sapien
8	89.5	8.9	859	Q8VDM6	Q8VDM6 mus musculus
9	89.5	8.9	3411	Q98803	Q98803 yellow feve
10	89.5	8.9	3411	Q91857	Q91857 yellow feve
11	89.5	8.9	3411	Q9YWN2	Q9YWN2 yellow feve
12	89.5	8.9	3411	Q9YWN1	Q9YWN1 yellow feve
13	89.5	8.9	3411	Q9YWN0	Q9YWN0 yellow feve
14	89.5	8.9	3411	Q89275	Q89275 yellow feve
15	89.5	8.9	3411	Q89277	Q89277 yellow feve
16	89.5	8.9	3411	Q89278	Q89278 yellow feve

17	89.5	8.9	3411	12	Q89276	Q89276 yellow feve
18	86.5	8.6	576	2	Q9RF25	Q9RF25 campylobact
19	86.5	8.6	3411	12	Q9YRV3	Q9YRV3 yellow feve
20	86	8.6	281	17	Q8TL57	Q8TL57 methanosarc
21	85.5	8.5	323	10	Q932W5	Q932W5 arabidopsis
22	84.5	8.4	576	2	Q85183	Q85183 campylobact
23	84	8.4	238	2	Q9FD07	Q9FD07 staphylococ
24	84	8.4	729	16	Q8YC77	Q8YC77 bruceella me
25	83.5	8.3	1004	15	Q994B3	Q994B3 human immun
26	83	8.3	263	16	Q8RCN9	Q8RCN9 fusbacteri
27	83	8.3	316	2	Q9FBG1	Q9FBG1 staphylococ
28	83	8.3	398	11	Q8VE75	Q8VE75 mus musculu
29	83	8.3	682	2	Q9EXB5	Q9EXB5 neisseria m
30	83	8.3	684	2	Q9JPM0	Q9JPM0 neisseria m
31	83	8.3	699	2	Q9EXC4	Q9EXC4 neisseria m
32	82.5	8.2	726	2	Q30713	Q30713 flavobacter
33	82.5	8.2	759	16	Q92A91	Q92A91 listeria in
34	82	8.2	402	16	Q9AB91	Q9AB91 caulobacter
35	82	8.2	798	10	Q9LLS0	Q9LLS0 nicotiana r
36	82	8.2	839	5	Q8T6J0	Q8T6J0 dictyosteli
37	81.5	8.1	409	12	Q96605	Q96605 avian infec
38	81.5	8.1	493	16	Q98PH9	Q98PH9 mycoplasma
39	81.5	8.1	549	10	Q93239	Q93239 arabidopsis
40	81.5	8.1	603	10	Q9SZT9	Q9SZT9 arabidopsis
41	81	8.1	239	2	Q9KH49	Q9KH49 staphylococ
42	81	8.1	239	16	Q53782	Q53782 staphylococ
43	81	8.1	283	2	Q9F474	Q9F474 ehrlichia c
44	81	8.1	663	16	Q8ZNL0	Q8ZNL0 salmonella
45	81	8.1	663	16	Q8Z597	Q8Z597 salmonella

ALIGNMENTS

RESULT 1

Q9PC94	PRELIMINARY;	PRT;	716 AA.
ID	Q9PC94		
AC	Q9PC94;		
DC	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)		
DE	Hypothetical protein Xf1887.		
GN	Xf1887.		
OS	Xylella fastidiosa.		
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;		
OC	Xylella.		
OX	NCBI_taxID=2371;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=9A5C;		
RX	MEDLINE=20365717; PubMed=10910347;		
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,		
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,		
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,		
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,		
RA	Colaudo N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,		
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,		
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,		
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,		
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,		
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,		
RA	Kriegel J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,		
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,		
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,		
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,		
RA	Menck C.F.M., Miraca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,		
RA	Moore D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,		
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,		
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,		
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,		
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,		
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,		
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,		

QY	65	GAWNYHTTGKGVLEKQDPKSDFAVOENILDLFRKYNTRYAENGQHLIAFLSNNDITG	120
Db	92	LAPGSALVTGQKYVAIGNPLGLTISEGLISLRDQDGRK-----SVQTSAAISR	144
QY	125	GNSGSPVFDKNGRLIGL	141
Db	145	GSSGGGLFDANGRLIGI	161
RESULT 3			
Q97LA8			
ID	Q97LA8	PRELIMINARY:	PRT: 513 AA.
AC	Q97LA8;		
DT	01-OCT-2001 (TrEMBLrel. 18, Created)		
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Sensory transduction histidine kinase.		
GN	CAC0654.		
OS	Clostridium acetobutylicum.		
OC	Bacteria; Firmicutes; Bacilli; Clostridium group; Clostridia;		
OC	Clostridiales; Clostridiaceae; Clostridium.		
OX	NCBI_TaxID=1488;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 824 / DSM 792 / VKM B-1787;		
RX	MEDLINE=21359325; PubMed=11466286;		
RA	Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,		
RA	Gibson R., Lee H.M., Dubois J., Qiu D., Hittl J., Wolf Y.I., Daly M.J.,		
RA	Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,		
RA	Bennett G.N., Koonin E.V., Smith D.R.		
RT	"Genome sequence and comparative analysis of the solvent-producing		
RT	bacterium Clostridium acetobutylicum."		
RL	J. Bacteriol. 183:4823-4838(2001).		
DR	EMBL; A5007581; AAK78631.1; .		
DR	InterPro: IPR003594; ATPbind_ATPase.		
DR	InterPro: IPR004358; Bact_sens_pr_C.		
DR	InterPro: IPR003660; HAMP.		
DR	InterPro: IPR003661; His_kinA.		
DR	InterPro: IPR004359; HIS_KIN_sig.		
DR	Pfam; PF00672; HAMP; 1.		
DR	Pfam; PF02518; HATPase_c; 1.		
DR	Pfam; PF00512; signal; 1.		
DR	PRINTS; PR00344; BCTRLSENSOR.		
DR	SMART; SM00387; HATPase_c; 1.		
DR	SMART; SM00388; HisKA; 1.		
KW	Kinase; Complete proteome.		
SQ	SEQUENCE 513 AA; 59185 MW; A2F61615182423B3 CRC64;		
Query Match 9.3%; Score 93; DB 16; Length 513;			
Best Local Similarity 24.7%; Pred. No. 3.9;			
Matches 37; Conservative 29; Mismatches 56; Indels 28; Gaps			
QY	48	FTWRMSYSGTKGYEPODGAWNYVHT-----TKGVLEKQDPKSDFAVOENILDLF	98
Db	96	FVYKKGVDVYKTEKRYQNKYTKQYLETFFNKGKVIITLTP--DDYS-EDNLSKCY	152
QY	99	RTKNGRYAENGQHLIAFLSNNDITGNSGSPVFDKNGRLIGL--AFDGNWEAMSGDI--	154
Db	153	GTKNL-KYYSKGQIFIKDLLKNMSG--KLSIVTDKNSVIRIQIKDDSVYDKMFGELSN	209
QY	155	-----EFEPDLQRTISVDIRYVLFMI	175
Db	210	NKGFKWTSGEFEAAVMHLYRRIYIIFII	239
RESULT 4			
Q96604			
ID	Q96604	PRELIMINARY:	PRT: 408 AA.
AC	Q96604;		
DT	01-FEB-1997 (TrEMBLrel. 02, Created)		
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	Nucleocapsid protein.		

```
OS Avian infectious bronchitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11120;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Q3/88;
RX MEDLINE=97124667; PubMed=8955062;
RA Sapats S.I., Ashton F., Wright P.J., Ignjatovic J.;
RT "Novel variation in the N protein of avian infectious bronchitis
RT virus.";
RL Virology 226:412-417(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Q3/88;
RA Sapats S.I., Ashton F., Wright P.J., Ignjatovic J.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U52600; AAB48161.1; -
DR InterPro; IPR001218; Corona_nucleocap.
DR Pfam; PF00937; Corona_nucleoca; 1.
SQ SEQUENCE 408 AA; 45426 MW; BCFC3ADAE75D13 CRC64;

Query Match          9.0%; Score 90.5; DB 12; Length 408;
Best Local Similarity 25.9%; Pred. No. 4.9;
Matches 37; Conservative 17; Mismatches 44; Indels 45; Gaps 7;

QY 25 GKRLFFAGFLRE-----MYPGRALPSDANFTMRMSYSGSIK-----GYEPQDGA 66
Db 10 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 32 QASWFSQSLAKKRTGEPFFEGSGVPDNSNVKPFQHGYYKRRYKSGGKRPVADA 91
Db 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 WYNYHT-TGK-GVLEKQDPKSDFAVOENILDLFRTKNYG-----RYAENGQL 112
Db 12 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 WYFYTTGTGPGDLEMDPNDVWVWAKAGADTSKIGNYGVDPDKFDQAPLRFIEGG-- 149
Db 13 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 113 HIAFLSN-----DITGNSGS 129
Db 14 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 -----PNNFRWDFIALSRGRNGS 167
Db 15 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
Q51284 PRELIMINARY; PRT; 711 AA.
ID Q51284
AC Q51284;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Transferrin-binding protein 2 precursor.
GN TBP2.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B163;
RA Legrain M., Findell A., Villevall D., Quentin-Millet M., Jacobs E.;
RT "Molecular characterization of hybrid transferrin-binding protein 2's
RT from Neisseria meningitidis.";
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z50731; CAA90598.1; -
DR InterPro; IPR001677; transferrin_bind.
DR Pfam; PF01298; Lipoprotein_5; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 20 POTENTIAL.
FT CHAIN 21 711 TRANSFERRIN-BINDING PROTEIN 2.
SQ SEQUENCE 711 AA; 77057 MW; 9BC8A1671F6991D0 CRC64;

Query Match          9.0%; Score 90; DB 2; Length 711;
Best Local Similarity 21.8%; Pred. No. 12;
Matches 38; Conservative 28; Mismatches 66; Indels 42; Gaps 7;

QY 2 KSVTAARATQADMANAYALEKRLFFAGLREMYPCRALPDSANTMRMSYSGSIKGYE 61
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 KNSKAMQAGESSQADAKTEQVGQSMFLQG--BRTDEKEIPSEQNIYVR----- 565
```

```
QY 62 PQCGAWNYN---HTTGKG-VLEKQDPKSDFAVOENILDLFRTKNYGRYAENGQLHIAFL 117
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 ---GSWYGHIASTSTWSGNSADKEGGNRAEFTV-----NFGEKKITGTL----- 606
QY 118 SNNDITGNSGSPVFDKNGRLIGLAFDGNWMSGDIEFEPDLQRTISVDIRYV 171
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 607 -----TAENRQEATFTIDGKIEGNFGSGT--AKTAELGFDLDQKNTTRTPKAYI 653
RESULT 6
O76022 PRELIMINARY; PRT; 856 AA.
ID O76022
AC O76022;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE EIB-55kDa-associated protein.
GN EIB-AP5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98406198; PubMed=9733834;
RA Gabler S., Schuett H., Groitl P., Wolf H., Shenk T., Dobner T.;
RT "EIB-55kDa-dalton-associated protein: A cellular protein with RNA-
RT binding activity implicated in nucleocytoplasmic transport of
RT adenovirus and cellular mRNAs.";
RL J. Virol. 72:7960-7971(1998).
DR EMBL; AJ007509; CAA07548.1; -
DR InterPro; IPR003034; SAP.
DR InterPro; IPR003878; SPRY domain.
DR InterPro; IPR003877; SPRY_receptor.
DR Pfam; PF02037; SAP; 1.
DR Pfam; PF00622; SPRY; 1.
DR SMART; SM00513; SAP; 1.
DR SMART; SM00449; SPRY; 1.
SQ SEQUENCE 856 AA; 95809 MW; 937D6ACD1BD45DFF CRC64;

Query Match          8.9%; Score 89.5; DB 4; Length 856;
Best Local Similarity 25.1%; Pred. No. 17;
Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;

QY 19 AYALEKCKRLFFAGLREMYPCRALPSDA--NFTMR-----SYGSIKGYEPQDGAWNYNH 71
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 249 SYGVRGRVCFEMKINEISVKHLPSTEPDPHVVRIGWSLDCSTQLGEEPFSS---YGYG 305
QY 72 TTGKGVLEKQDPKSDFAVOENILDLFRTKNYG-RYAENGQLHIAFLSNNDITG----- 124
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 GTGK-----KSTN-----SRFENYGDKFAE-----NDVICGFADFE 336
QY 125 -GNSGSPVFDKNGRLIGLAFDGNWMSGDIEFEPDLQRTISVDIRY 170
Db 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 337 CGNDVELSFTKNGKWMGIAFRIQKEALGGQALYPHVLVKNCAVEFN 383
RESULT 7
Q9BUJ2 PRELIMINARY; PRT; 856 AA.
ID Q9BUJ2
AC Q9BUJ2; Q9UG75;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE EIB-55kDa-associated protein 5.
GN DKFZP586D0920.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```



Query Match 8.9%; Score 89.5; DB 12; Length 3411;  
Best Local Similarity 25.5%; Pred. No. 1.1e+02;  
Matches 39; Conservative 21; Mismatches 46; Indels 47;

```

DR PROSITE; PS00962; RIBOSOMAL_S2.1; UNKNOWN.1.
KW ATP-binding; Helicase.
SQ SEQUENCE 3411 AA; 379556 MW; 1AA3ED795108EABC CRC64;

Query Match 8.98; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

Qy 19 AYATEKGRKRLFFAGLRMPGRALPSDANETMRM--SYGSIKGYEPQDGAWNYH----- 71
Db 1542 AFLVRNKKL-----IPSWASVVEDLVAYG---GSWKLEGRWDGEEVEQVILIA 1585

Qy 72 -TTGKGVLEKODPKSDEFAVOENITDLFTKTKNYGAFNGOLHIAFLSNNDITGNSGSP 130
Db 1586 AVPKGNVNVQTKFS-----LFKVRNGG--EIGAVAL-----DYPSTGTSOSP 1625

Qy 131 VFDKNRGLRGLIAGFDGNWEAMSGDIEFEFDLQRT 163
Db 1626 IIVNRGEVIGLYGNG--ILVGDNSFVSASISQT 1655

RESULT 13
Q9YWN0
ID Q9YWN0 PRELIMINARY; PRT; 3411 AA.
AC Q9YWN0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
DE Yellow fever virus.
OS Yellow fever virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17D-204-SOUTH AFRICA VACCINE;
RX MEDLINE=98378039; Pubmed=9714237;
RT Xie H., Ryman K.D., Campbell G.A., Barrett A.D.T.;
RT "Mutation in NS5 protein attenuates mouse neurovirulence of yellow
RT fever 17D vaccine virus."
RL J. Gen. Virol. 79:1895-1899(1998).
RL EMBL; AF052446; AAC35908.1;
RS HSP; P14336; 1SVB.
DR MEROPS: S07.001;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001122; Flavi_capsidC.
DR InterPro; IPR000336; Flavi_glycoprote.
DR InterPro; IPR001850; Flavi_helicase.
DR InterPro; IPR000069; Flavi_M.
DR InterPro; IPR001157; Flavi_NS1.
DR InterPro; IPR000752; Flavi_NS2A.
DR InterPro; IPR000487; Flavi_NS2B.
DR InterPro; IPR000404; Flavi_NS4A.
DR InterPro; IPR001528; Flavi_NS4B.
DR InterPro; IPR00208; Flavi_NS5.
DR InterPro; IPR002535; Flavi_propep.
DR InterPro; IPR002877; FtsJ.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001865; Ribosomal_S2.
DR Pfam; PF01003; Flavi_capsid; 1.
DR Pfam; PF00869; Flavi_glycoprot; 1.
DR Pfam; PF02832; Flavi_glycop_C; 1.
DR Pfam; PF00949; Flavi_helicase; 1.
DR Pfam; PF01004; Flavi_M; 1.
DR Pfam; PF00948; Flavi_NS1; 1.
DR Pfam; PF01005; Flavi_NS2A; 1.
DR Pfam; PF01002; Flavi_NS2B; 1.
DR Pfam; PF01350; Flavi_NS4A; 1.
DR Pfam; PF01349; Flavi_NS4B; 1.
DR Pfam; PF00972; Flavi_NS5; 1.
DR Pfam; PF01570; Flavi_propep; 1.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF00271; helicase_C; 1.

```



```
DR PRODOM: PD001496; Flavi_NS1; 1.
DR PRODOM: PD001536; Flavi_glycoprote; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 3411 AA; 379500 MW; 8A9B127F6623A933 CRC64;

Query Match      8.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

QY 19 AYAEKGRLEFFAGLRMYPGRALPSDANFTMRM-SYGSIKGYEPQDGAWNYH----- 71
Db 1542 AFLVRNGKKL-----IPSWASVKEDLVAYG---GSWKLEGRWDGEEVQLIA 1585

QY 72 -TTCKGVLEKQDKPSDFEAVQENILDFRTKNYGRYAENGOLHIAFLSNNDITGNSGSP 130
Db 1586 AVPKNNVNVQTKPS-----LFKVRNGG---EIGAVAL-----DYPSTGSGSP 1625

QY 131 VFDKNGRLIGLAFDGNWEAMSGDIEFEPDLQRT 163
Db 1626 IVNRNGEIVIGLYNG---ILVGDNSFVSAISQT 1655

RESULT 14
Q89275 ID Q89275 PRELIMINARY; PRT; 3411 AA.
AC Q89275;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Polypeptin.
OS Yellow fever virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17DD;
RA dos Santos C.N., Post P.R., Carvalho R., Ferreira I.I., Rice C.M.,
  Galler R.;
RT "Complete nucleotide sequence of yellow fever virus vaccine strains
  17DD and 17B-213.";
RL Virus Res. 35:35-41(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17DD;
RA Galler R.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U17066; AAC54267.1; -.
DR HSSP; P14336; ISVB.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001122; Flavi_capsidC.
DR InterPro; IPR000336; Flavi_glycoproteE.
DR InterPro; IPR001850; Flavi_helicase.
DR InterPro; IPR000069; Flavi_M.
DR InterPro; IPR001157; Flavi_NS1.
DR InterPro; IPR000752; Flavi_NS2A.
DR InterPro; IPR000487; Flavi_NS2B.
DR InterPro; IPR000404; Flavi_NS4A.
DR InterPro; IPR001528; Flavi_NS4B.
DR InterPro; IPR000208; Flavi_NS5.
DR InterPro; IPR002535; Flavi_propep.
DR InterPro; IPR002877; FtsJ.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001863; Ribosomal_S2.
DR Pfam; PF01003; Flavi_capsid; 2.
DR Pfam; PF00869; Flavi_glycoprot; 2.
DR Pfam; PF02832; Flavi_glycop_C; 2.
DR Pfam; PF00949; Flavi_helicase; 2.
DR Pfam; PF01004; Flavi_M; 2.
DR Pfam; PF00948; Flavi_NS1; 2.
```

```
DR Pfam; PF01005; Flavi_NS2A; 2.
DR Pfam; PF01002; Flavi_NS2B; 2.
DR Pfam; PF01350; Flavi_NS4A; 2.
DR Pfam; PF01349; Flavi_NS4B; 2.
DR Pfam; PF00972; Flavi_NS5; 2.
DR Pfam; PF01570; Flavi_propep; 2.
DR Pfam; PF01728; FtsJ; 2.
DR Pfam; PF00271; Helicase_C; 2.
DR PRODOM: PD001496; Flavi_NS1; 1.
DR PRODOM: PD001536; Flavi_glycoprote; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 2.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_2.
KW ATP-binding; Helicase.
FT CHAIN 2 121 CAPSID PROTEIN.
FT CHAIN 211 285 M PROTEIN.
FT CHAIN 286 778 ENVELOPE PROTEIN.
FT CHAIN 779 1187 NS1.
FT CHAIN 1188 1354 POTENTIAL.
FT CHAIN 1355 1483 POTENTIAL.
FT CHAIN 1484 2106 NS3.
FT CHAIN 2108 2394 POTENTIAL.
FT CHAIN 2395 2506 POTENTIAL.
FT CHAIN 2507 2507 NS5.
SQ SEQUENCE 3411 AA; 379429 MW; 0E7C8D189524790B CRC64;

Query Match      8.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

QY 19 AYAEKGRLEFFAGLRMYPGRALPSDANFTMRM-SYGSIKGYEPQDGAWNYH----- 71
Db 1542 AFLVRNGKKL-----IPSWASVKEDLVAYG---GSWKLEGRWDGEEVQLIA 1585

QY 72 -TTCKGVLEKQDKPSDFEAVQENILDFRTKNYGRYAENGOLHIAFLSNNDITGNSGSP 130
Db 1586 AVPKNNVNVQTKPS-----LFKVRNGG---EIGAVAL-----DYPSTGSGSP 1625

QY 131 VFDKNGRLIGLAFDGNWEAMSGDIEFEPDLQRT 163
Db 1626 IVNRNGEIVIGLYNG---ILVGDNSFVSAISQT 1655

RESULT 15
Q89277 ID Q89277 PRELIMINARY; PRT; 3411 AA.
AC Q89277;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Polypeptin.
OS Yellow fever virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRENCH NEUROTROPIC VIRUS;
MEDLINE=96068808; PubMed=7595382;
RA Wang E., Ryman K.D., Jennings A.D., Wood D.J., Taffs F., Minor P.D.,
  Sanders P.G., Barrett A.D.;
RT "Comparison of the genomes of the wild-type French viscerotropic
  strain of yellow fever virus with its vaccine derivative French
  neurotropic vaccine.";
RL J. Gen. Virol. 76:2749-2755(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRENCH NEUROTROPIC VIRUS;
RA Wang E.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U21055; AAA99712.1; -.
DR HSSP; P14336; ISVB.
DR InterPro; IPR001410; DEAD.
```

```

DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000336; Flavi_glycoproteE.
DR InterPro: IPR001850; Flavi_helicase.
DR InterPro: IPR000069; Flavi_M.
DR InterPro: IPR001157; Flavi_NS1.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR002535; Flavi_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001865; Ribosomal_S2.
DR Pfam: PF01003; Flavi_capsid; 2.
DR Pfam: PF00889; Flavi_glycoprot; 2.
DR Pfam: PF02832; Flavi_glycop_C; 2.
DR Pfam: PF00949; Flavi_helicase; 2.
DR Pfam: PF01004; Flavi_M; 2.
DR Pfam: PF00948; Flavi_NS1; 2.
DR Pfam: PF01005; Flavi_NS2A; 2.
DR Pfam: PF01002; Flavi_NS2B; 2.
DR Pfam: PF01350; Flavi_NS4A; 2.
DR Pfam: PF01349; Flavi_NS4B; 2.
DR Pfam: PF00972; Flavi_NS5; 2.
DR Pfam: PF01570; Flavi_propep; 2.
DR Pfam: PF01728; FtsJ; 2.
DR Pfam: PF00271; helicase_C; 2.
DR ProDom: PD001496; Flavi_NS1; 1.
DR ProDom: PD001556; Flavi_glycoproteE; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELICC; 2.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_2.
KW ATP-binding; Helicase.
FT CHAIN 2 121 CAPSID PROTEIN C.
FT CHAIN 211 285 M PROTEIN.
FT CHAIN 286 778 ENVELOPE PROTEIN E.
FT CHAIN 779 1187 NS1.
FT CHAIN 1188 1354 NS2A.
FT CHAIN 1355 1483 NS2B.
FT CHAIN 1484 2106 NS3.
FT CHAIN 2108 2394 NS4A.
FT CHAIN 2395 2506 NS4B.
FT CHAIN 2507 3411 NS5.
SQ SEQUENCE 3411 AA; 379183 MW; E7A5E79C999C9D8D CRC64;

Query Match 8.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

Qy 19 AYATEKGRLLFFAGLRMYPGRALPSPDANFTMRM-SYGSIKGYEPDGDGAWNYNH----- 71
Db 1542 AFLVRNGKRL-----IPSNASVKEDLVAYG---GSWKLEGRWDGEEVQLIA 1585

Qy 72 -TTGKGVLKQDPKSDFAVGQENILDFRTKNRYAENGQHLIAFLSNNDITGNSGSP 130
Db 1586 AVPGKNVNVVQTKPS-----LFKVRNGG---EIGAVAL-----DYPSTGSGSP 1625

Qy 131 VFDKNGRLIGLAFDGNWAMSGDIEFPDLQRT 163
Db 1626 IVNRNGEVIIGLYGNG---ILVGDNSFVSAISQT 1655

```

Search completed: May 25, 2003, 15:36:18  
Job time : 26.1716 secs

Result No.	Query Match	Score	Length	DB	ID	Description	
1	95	9.5	507	1	YY42_CAEEL	Q18416	caenorhabdi
2	89.5	8.9	3411	1	POLG_YEFV1	P03314	y genome po
3	89.5	8.9	3411	1	POLG_YEFV2	P19901	y genome po
4	86.5	8.6	575	1	FLB2_CAMJ2	P22252	campylobact
5	86	8.6	1060	1	DP3A_LACLA	Q9c170	lactococcus
6	85.5	8.5	321	1	HH0A_ARATH	O9sel7	arabidopsis
7	84.5	8.4	256	1	TRYE_DROER	P54627	drosophila
8	83	8.3	698	1	TBFB_NEIMA	O68937	neisseria m
9	82	8.2	504	1	VPS_BPMU	O9rtiv0	bacterioph
10	82	8.2	974	1	BCA1_MOUSE	Q61140	mus musculu
11	81	8.1	968	1	BCA1_RAT	Q63767	rattus norv
12	80.5	8.0	433	1	MCRB_METBA	P07955	methanosarc
13	80	8.0	336	1	STSP_STAAU	P04188	staphylococ
14	80	8.0	3433	1	POLG_KUNJM	P14335	k genome po
15	79.5	7.9	747	1	Y030_UREPA	O9prb5	ureaplasma
16	78.5	7.8	1683	1	POLG_DEN2T	P27914	dengue viru
17	78.5	7.8	3388	1	POLG_DEN2P	P12823	d genome po
18	78	7.8	870	1	BCA1_HUMAN	P56945	homo sapien
19	78	7.8	3432	1	POLG_TAEVJ	P32886	j genome po
20	77.5	7.7	971	1	CLA4_CANAL	O14427	candidate alb
21	77.5	7.7	3391	1	POLG_DEN2T	P29991	d genome po
22	77	7.7	686	1	EMAP_STRPU	Q26613	strongyloce
23	77	7.7	3430	1	POLG_WNV	P06935	w genome po
24	76.5	7.6	256	1	TRYE_DROME	P35005	drosophila
25	76.5	7.6	458	1	APFA_BORBU	Q45055	borrelia bu
26	76.5	7.6	647	1	CN16_ECOLI	P08331	escherichia
27	76.5	7.6	676	1	HSF2_TRYBB	P20030	trypanosoma
28	76	7.6	521	1	Y92_MYCPN	O50335	mycoplasma
29	76	7.6	560	1	5NTD_VIBPA:	P22848	vibrio para
30	76	7.6	647	1	CN16_SALTY	P26265	salmonella
31	75.5	7.5	434	1	ENO_DBSVM	O32513	desulfovibr
32	75.5	7.5	787	1	RELA_MYCLE	Q49640	mycobacteri
33	75.5	7.5	790	1	RELA_MYCTU	Q50638	mycobacteri

POLG\_YEFV1 STANDARD; PRT; 3411 AA.

AC P03314; 042028;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)].

OS Yellow fever virus (strain 17D).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Flavivirus.

OX NCBI\_TaxID=11090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=6527570; PubMed=4023707;

RA Rice C.M., Lencches E.M., Eddy S.R., Shin S.J., Sheets R.L., Straus J.H.;

RT "Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution.";

RL Science 229:726-733(1995).

CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.

CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + [RNA](N).

CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X03700; CAA27332.1; -

DR PIR; A03914; GNMVY.

DR HSSP; P14336; 1SVB.

DR MEROPS; S07.001; -

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR000069; Flavi\_M.

DR InterPro; IPR001157; Flavi\_NS1.

DR InterPro; IPR000752; Flavi\_NS2A.

DR InterPro; IPR000487; Flavi\_NS2B.

DR InterPro; IPR000404; Flavi\_NS4A.

DR InterPro; IPR001528; Flavi\_NS4B.

DR InterPro; IPR000208; Flavi\_NS5.

DR InterPro; IPR001122; Flavi\_capsidC.

DR InterPro; IPR000336; Flavi\_glycoproteE.

DR InterPro; IPR001850; Flavi\_helicase.

DR InterPro; IPR002535; Flavi\_propep.

DR InterPro; IPR002877; FtsJ.

DR InterPro; IPR001650; Helicase\_C.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00869; Flavi\_glycoprot; 1.

DR Pfam; PF00948; Flavi\_NS1; 1.

DR Pfam; PF00949; Flavi\_helicase; 1.

DR Pfam; PF00972; Flavi\_NS5; 1.

DR Pfam; PF01002; Flavi\_NS2B; 1.

DR Pfam; PF01003; Flavi\_capsid; 1.

DR Pfam; PF01004; Flavi\_M; 1.

DR Pfam; PF01005; Flavi\_NS2A; 1.

DR Pfam; PF01349; Flavi\_NS4B; 1.

DR Pfam; PF01350; Flavi\_NS4A; 1.

DR Pfam; PF01570; Flavi\_propep; 1.

DR Pfam; PF01728; FtsJ; 1.

DR Pfam; PF02832; Flavi\_glycop\_C; 1.

DR ProDom; PD001496; Flavi\_NS1; 1.

DR ProDom; PD001556; Flavi\_glycoproteE; 1.

DR SMART; SM00490; HelicC; 1.

KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Hydrolyase; Helicase;

KW ATP-binding; Transmembrane; Nonstructural protein.

FT INIT\_MET 1 1

FT CHAIN 1 121

FT PROPEP 122 210

FT CHAIN 211 285

FT CHAIN 286 778

FT CHAIN 779 1130

FT CHAIN 1131 1354

FT CHAIN 1355 1484

FT CHAIN 1485 2107

FT CHAIN 2108 2256

FT CHAIN 2257 2506

FT CHAIN 2507 3411

FT TRANSEM 249 269

FT TRANSEM 271 285

FT TRANSEM 740 753

FT TRANSEM 755 778

FT TRANSEM 1159 1180

FT DOMAIN 383 396

FT NP\_BIND 1682 1689

FT SITE 1773 1776

FT DISULFID 288 315

FT DISULFID 345 401

FT DISULFID 359 390

FT DISULFID 377 406

FT DISULFID 467 568

FT DISULFID 585 615

FT CARBOHYD 134 134

FT CARBOHYD 150 150

FT CARBOHYD 908 908

FT CARBOHYD 986 986

FT CARBOHYD 2320 2320

FT CARBOHYD 2346 2346

FT CARBOHYD 2467 2467

FT CARBOHYD 3411 AA; 379512 MW; 680E0FACD23DFA6 CRC64;

SQ SEQUENCE 3411 AA; 379512 MW; 680E0FACD23DFA6 CRC64;

Query Match 8.98; Score 89.5; DB 1; Length 3411;

Best Local Similarity 25.58; Pred. No. 10;

Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

Qy 19 AYATEKGRLEFFAGLREMYPGRALPSDANFTWRM--SYGSIKGYEPQDGAWNYH----- 71

Db 1542 AFLVRNCKKL-----IPSWASVKEDLVAVG--GSMKLEGRWDGEEVQLIA 1585

Qy 72 -TTGKGVLEKQDKPSDEFAVOENILDFRTKNYGRYAENGQIHAFLSNNDITGNSGSP 130

Db 1586 AVPGKNVNVNQTGPS-----LfkVRNGG---EIGAVAL-----DYPSTGSGSP 1625

Qy 131 VFDKNGRLIGLAFDGNWEAMSCDTEFEFDLQRT 163

Db 1626 IVNRNGEVIGLYGNG---ILVGDNSFVSATISQT 1655

RESULT 3

POLG\_YEFV2 STANDARD; PRT; 3411 AA.

AC P19901;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)].

(NS5)]. fever virus (strain Pasteur 17D-204).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 OC NCBI\_TaxID=11091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9292413; PubMed=2734112;  
 RA Dupuy A., Despres P., Cahour A., Girard M., Bouloy M.;  
 RT "Nucleotide sequence comparison of the genome of two 17D-204 yellow  
 fever vaccines";  
 RL Nucleic Acids Res. 17:3989-3989(1989).  
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein, commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 CC PROTEIN C AND RNA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X15062; CAB37419.1; -.  
 DR PIR; S07757; GNVYP.  
 DR HSSP; P14336; 1SVB.  
 DR MEROPS; S07.001; -.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR001157; Flavi\_NS1.  
 DR InterPro; IPR000752; Flavi\_NS2A.  
 DR InterPro; IPR000487; Flavi\_NS2B.  
 DR InterPro; IPR000404; Flavi\_NS4B.  
 DR InterPro; IPR001528; Flavi\_NS4B.  
 DR InterPro; IPR000208; Flavi\_NS5.  
 DR InterPro; IPR001122; Flavi\_capsidC.  
 DR InterPro; IPR000336; Flavi\_glycoproteE.  
 DR InterPro; IPR001850; Flavi\_helicase.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR InterPro; IPR002877; FtsJ.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprote; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF00949; Flavi\_helicase; 1.  
 DR Pfam; PF00972; Flavi\_NS5; 1.  
 DR Pfam; PF01002; Flavi\_NS2B; 1.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF01005; Flavi\_NS2A; 1.  
 DR Pfam; PF01349; Flavi\_NS4B; 1.  
 DR Pfam; PF01350; Flavi\_NS4A; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR Pfam; PF01728; FtsJ; 1.  
 DR Pfam; PF02832; Flavi\_glycop-C; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR ProDom; PD001556; Flavi\_glycoproteE; 1.  
 DR SMART; SM00490; HELIC\_C; 1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;  
 KW ATP-binding; Transmembrane; Nonstructural protein.  
 FT INT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 CC CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 121 CAPSID PROTEIN C.  
 FT PROPEP 122 210 ENVELOPE GLYCOPROTEIN M.  
 FT CHAIN 211 285 MAJOR ENVELOPE PROTEIN E.  
 FT CHAIN 286 778 NONSTRUCTURAL PROTEIN NS1.  
 FT CHAIN 779 1130 NONSTRUCTURAL PROTEIN NS2A.  
 FT CHAIN 1131 1354 NONSTRUCTURAL PROTEIN NS2B.  
 FT CHAIN 1355 1484 PROTEASE/HELICASE (NS3).  
 FT CHAIN 1485 2107 NONSTRUCTURAL PROTEIN NS4A.  
 FT CHAIN 2108 2256 NONSTRUCTURAL PROTEIN NS4B.  
 FT CHAIN 2257 2506 RNA-DIRECTED RNA POLYMERASE (NS5).  
 FT CHAIN 2507 3411 ATP (POTENTIAL).  
 FT NP\_BIND 1682 1689 DEAH BOX.  
 FT SITE 1773 1776 POTENTIAL.  
 FT TRANSMEM 249 269 POTENTIAL.  
 FT TRANSMEM 271 285 POTENTIAL.  
 FT TRANSMEM 740 753 POTENTIAL.  
 FT TRANSMEM 755 778 POTENTIAL.  
 FT TRANSMEM 1159 1180 POTENTIAL.  
 FT DISULFID 288 315 BY SIMILARITY.  
 FT DISULFID 345 401 BY SIMILARITY.  
 FT DISULFID 359 390 BY SIMILARITY.  
 FT DISULFID 377 406 BY SIMILARITY.  
 FT DISULFID 467 568 BY SIMILARITY.  
 FT DISULFID 585 615 BY SIMILARITY.  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 150 150 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 908 908 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 986 986 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2320 2320 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2346 2346 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2467 2467 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 3411 AA; 379524 MW; 3298C0771FED23F7 CRC64;  
 Query Match 8.9%; Score 89.5; DB 1; Length 3411;  
 Best Local Similarity 25.5%; Pred. No. 10;  
 Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;  
 QY 19 AYAIEKGRLEFAGLRMPGRALPSDANFTMRM-SYGSIKGYEPQDGAWNYH----- 71  
 Db 1542 AFLVRNGKKL-----IPSWASVKEDLVAYG---GSWKLEGRWDGEERVLIA 1585  
 QY 72 -TTGKGVLEKODPKSDEFAVOENILDLPRTKNYGRYAENGQLHTAFLSNNDITGNSGSP 130  
 Db 1586 AVPGKNVNVQTKPS-----LFKVRNGG---EIGAVAL-----DYPSTGSGSP 1625  
 QY 131 VFDKNGRLIGLAFDGNWEAMSGDIEFEDLORT 163  
 Db 1626 IVNRNGEVIGLYNG---ILVGDNSFVSAISQT 1655  
 RESULT 4  
 ID\_FLB2\_CAMJE STANDARD; PRT; 575 AA.  
 AC P22252;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Flagellin B.  
 GN FLAB.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=81116;  
 RX MEDLINE=91009243; PubMed=2211662;  
 RA Nuijten P.J., van Asten F.J., Gastra W., van der Zeijst B.A.;  
 RT "Structural and functional analysis of two Campylobacter jejuni  
 flagellin genes";  
 RL J. Biol. Chem. 265:17798-17804(1990).  
 CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
 CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.

CC -!- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J05635; AAA23025.1; -  
 CC PIR: B39228; B39228.  
 CC InterPro: IPR001492; FlagellinN.  
 CC InterPro: IPR001029; Flagellin\_C.  
 CC Pfam: PF00669; Flagellin\_N: 1.  
 CC Pfam: PF00700; Flagellin\_C: 1.  
 CC PRINTS: PR00207; FLAGELLIN.  
 CC ProDom: PD000316; Flagellin\_C: 1.  
 CC Flagella.  
 CC FT INIT\_MET 0 0 BY SIMILARITY.  
 CC SEQUENCE 575 AA; 59728 MW; D0531AF308A7BF1D CRC64;  
 SQ  
 Query Match 8.6%; Score 86.5; DB 1; Length 575;  
 Best Local Similarity 31.5%; Pred. No. 2.2;  
 Matches 28; Conservative 15; Mismatches 39; Indels 7; Gaps 3;  
 QY 66 AWNYHTTKGVLEKQDPKSFQVQENILDLFRPKNGRYAENGOLHIAFLSNNDITGG 125  
 DB 231 ATYDKVTKTGVVAT-KEGTSQDFALNGVVGQI---NYKDGNNQQLVSAINAVKDTTGV 286  
 QY 136 NSGSPVFDKNGRLGLAFDGNWAMSGDI 154  
 DB 287 QASK---DENGKLVTSADGRGIKTGDI 312  
 QY 126 NSGSPVFDKNGRLGLAFDGNWAMSGDI 154  
 DB 287 QASK---DENGKLVTSADGRGIKTGDI 312  
 RESULT 5  
 ID DP3A\_LACLA STANDARD; PRT; 1060 AA.  
 AC Q9C170;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA polymerase III alpha subunit (EC 2.7.7.7).  
 GN DNAE OR LI0496.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=21235186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Mauer S., Jaillon O., Malarne K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403.";  
 RL Genome Res. 11:731-753(2001).  
 CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME  
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.  
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.  
 CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + [DNA](N).  
 CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA  
 CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE  
 CC DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH  
 CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,  
 CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE006285; AAK04594.1; -  
 CC InterPro: IPR003141; PHP\_N.  
 CC InterPro: IPR004805; PolC\_alpha.  
 CC InterPro: IPR004365; trna\_anti.  
 CC Pfam: PF01336; trna\_anti; 1.  
 CC Pfam: PF02231; PHP\_N; 1.  
 CC SMART: SM00481; POLIICAC; 1.  
 CC TIGRFAMS: TIGR00594; polc; 1.  
 KW Transferase; DNA-directed DNA polymerase; DNA replication;  
 KW Complete proteome.  
 SQ SEQUENCE 1060 AA; 121571 MW; 44B03F0937A8D3E6 CRC64;  
 Query Match 8.6%; Score 86; DB 1; Length 1060;  
 Best Local Similarity 21.7%; Pred. No. 5.1;  
 Matches 45; Conservative 31; Mismatches 73; Indels 58; Gaps 8;  
 QY 14 DAMANAYAIKGRKLFAGLRMPGRALPSDANPT----MRMSYGSITKGYEPQDQAWYN 69  
 DB 743 DALENGFEIEK-----PSINLMKIGDFVKNKKIRLGLAHVQGISRDILAKWTV 789  
 QY 70 YHTTCKGVLEKQDPKSFQVQENILDL-----FRPKNGRYAENGOLHIAFLS--NN 120  
 DB 790 ENQPYKGLADFEKLEPNFHHKKNILPLIQICAFDYSNRKGLAYNLADHANLJNYSD 849  
 QY 121 DITGNSGSPV-----FDKNGRLGLA-----FDGNWAMSGD 153  
 DB 850 DIFWASSGGGAYHEAEDYSETEKYDFEKN--LLGIVTTPHPLQNLARRFEGNFTPLA-- 905  
 QY 154 IEFEPDLQRTISVDIRYVLFMDIKWQ 180  
 DB 906 -QLVKNRRMTILVEINYIRTHRTKTGQ 931  
 RESULT 6  
 ID HHOA\_ARATH STANDARD; PRT; 321 AA.  
 AC Q9SEL7; Q49507;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Protease HhoA, chloroplast precursor (EC 3.4.21.-).  
 GN HHOA OR AT4G18370 OR F28J12.30.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC LENSCH M.H.A., Sokolenko A., Hermann R.G.;  
 RX "Identification and characterization of the chloroplast HhoA protease,  
 RX a homolog to the bacterial periplasmic protease HhoA.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,  
 RA Harris B., Ansorge W., Brandt P., Grivell L., Rieger M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,  
 RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren J., Glynoprez B., Chuang Y.-J., Vandenbussche F.,

RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,  
 RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buyschaert C., Giesen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,  
 RA Pettett A., Rajandream M.A., Lyne M., Benes V., Lechmann S.,  
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
 RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,  
 RA Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Lechman A., Aubourg S.,  
 RA Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Shkon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshua C.,  
 RA Antonio B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Grant S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
 RA Chen E., Marra M., Martienssen R., McCombie W.R.,  
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RT thaliana".  
 RL Nature 402:769-777(1999).  
 RN [3].  
 RP SEQUENCE OF 72-82; 96-110; 150-159; 178-211 AND 306-320.  
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,  
 RA Kieselbach T.,  
 RT "The chloroplast lumen from Arabidopsis thaliana".  
 RL Submitted (JUL-2001) to the SWISS-PROT data bank.  
 CC -1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 52C.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A GENE  
 CC PREDICTION ERROR AND THE FUSION OF THE CDS FOR THIS PROTEIN WITH  
 CC THE CDS FOR A PUTATIVE NUCLEIC ACID BINDING PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF114386; AAF24060.1;  
 CC EMBL; AL021710; CAA16717.1; ALT\_SEQ.  
 CC EMBL; AL161548; CAB78839.1; ALT\_SEQ.  
 CC MEROPS; S01.279;  
 CC InterPro; IPR001940; Protease2C.  
 CC InterPro; IPR001254; Ser\_protease\_Try.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00834; PROTEASES2C.  
 KW Hydrolase; Serine protease; Chloroplast; Thylakoid; Transit peptide.  
 FT TRANSIT 1 26  
 FT CHAIN 27 71 CHLOROPLAST (POTENTIAL).  
 FT DOMAIN 72 321 PROTEASE HHOA.  
 FT POLY-GLU 77 87  
 FT ACT\_SITE 145 145 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 186 186 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 264 264 CHARGE RELAY SYSTEM (POTENTIAL).

FT CONFLICT 40 40 R -> G (IN REF. 1).  
 SQ SEQUENCE 321 AA; 34691 MW; 68DB81E0BD27A7A7 CRC64;  
 Query Match 8.5%; Score 85.5; DB 1; Length 321;  
 Match Local Similarity 27.6%; Pred.No.1.3;  
 Matches 42; Conservative 16; Mismatches 67; Indels 27; Gaps 8;  
 QY 45 DANFTMRMSYSGSIKGYEP-QDGAWNYNHTTKG-----GVLEKQDPKSDFAV-----Q 91  
 DB 166 DAKTRFSKEGKIVGLDPDNDLAVLKTETEGRELNPVVLGTNDLRVQSCFAIGNPYGY 225  
 QY 92 ENILDLRTKNYGRY--AENCO-LHIAFLSNNDITGGNSGSPVFDKNGRLIGL---AFDG 145  
 DB 226 ENTUIGVWSGLGREIPSPNGKSISEAIQTADINSGSGPLDLSYGHHTGVNTATFTR 285  
 QY 146 NWEAMSGDIEPEPLDQRTISVD--IRVLFMI 175  
 DB 286 KSGMSSGYNF-----AIPDTVTRVPYLI 311  
 RESULT 7  
 TRYE\_DROER  
 ID TRYE\_DROER STANDARD; PRT; 256 AA.  
 AC PS4627;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trypsin epsilon precursor (EC 3.4.21.4).  
 GN TRY-EPISILON  
 OS Drosophila erecta (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7220;  
 RN [1]  
 RP SEQUENCE FROM N.A.A.  
 RA Wang S., Hickey D.A.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBDJ databases.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 51.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U40653; AAA83240.1;  
 CC HSP; P00763; 1DPO.  
 CC MEROPS; S01.112;  
 CC FlyBase; FBgn0015080; DereTry-epsilon.  
 CC InterPro; IPR001314; Chymotrypsin.  
 CC InterPro; IPR001254; Ser\_protease\_Try.  
 CC Pfam; PF00089; trypsin; 1.  
 CC PRINTS; PR00722; CHYMOTRYPSIN.  
 CC SMART; SM00020; Tryp-SPC; 1.  
 CC PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 CC PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 CC PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Serine protease; Zymogen; Signal; Multigene family.  
 FT SIGNAL 1 22 PROBABLE.  
 FT PROPEP 23 30 ACTIVATION PEPTIDE.  
 FT CHAIN 31 256 TRYPSIN EPSILON.  
 FT ACT\_SITE 71 71 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 56 72 BY SIMILARITY.  
 FT DISULFID 180 197 BY SIMILARITY.  
 FT DISULFID 206 230 BY SIMILARITY.  
 FT DISULFID 204 204 REQUIRED FOR SPECIFICITY (BY SIMILARITY).

[illegible]



Db 315 LRP---LWNNATGAVSMRGLNVSGDGLSDRFAINSNGMWIOMRONNAIFGKNIVNTD 371

Qy 97 ----LFRTKNYGRY-----AENGQLHIAFLSNNDITGNSGSPVDFKNKR----- 137

Db 372 SAQALLRQNHADRFWMIGGLGNKQFGIYMINNSRTANGDQAYMDNNGNLGCAQVLP 431

Qy 138 -----LIGLAFDGNWEAMSGD 153

Db 432 NYANFDSRYVRDLRGTLQSLAGGLSRDYKAPSGHVITGFTHTNGDWEMOGGD 482

RESULT 10

BCAL\_MOUSE STANDARD; PRT; 874 AA.

AC Q61140; Q60869;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE CRK-associated substrate (p130CAS) (Breast cancer anti-estrogen resistance 1 protein).

GN BCAR1 OR CRKAS OR CAS.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

[1]

RN SEQUENCE FROM N.A. (ISOFORMS CAS-A AND CAS-B), AND INTERACTION WITH FOCAL ADHESION KINASE.

RP TISSUE=Embryo;

RX MEDLINE=96068679; PubMed=7479864;

RA polte T.R., Hanks S.K.;

RT "Interaction between focal adhesion kinase and Crk-associated tyrosine kinase substrate p130Cas."

RL Proc. Natl. Acad. Sci. U.S.A. 92:10678-10682(1995).

[2]

RN INTERACTION WITH NEPHROCISTIN.

RA MEDLINE=20249316; PubMed=10739664;

RA Donaldson J.C., Dempsey P.J., Reddy S., Bouton A.H., Coffey R.J., Hanks S.K.;

RT "Crk-associated substrate p130(Cas) interacts with nephrrocystin and both proteins localize to cell-cell contacts of polarized epithelial cells";

RL Exp. Cell Res. 256:168-178(2000).

CC -!- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION. IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY). HAS BEEN SHOWN TO BE ESSENTIAL IN CARDIOVASCULAR DEVELOPMENT DURING EMBRYOGENESIS.

CC -!- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1, ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL (BY SIMILARITY). INTERACTS WITH NEPHROCISTIN AND PTK2B.

CC -!- SUBCELLULAR LOCATION: FOCAL ADHESIONS. UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO THE MEMBRANE UPON TYROSINE PHOSPHORYLATION (BY SIMILARITY).

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: CAS-B (SHOWN HERE) AND CAS-A; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-BINDING SITES POTENTIALLY BIND CRK, NCK AND ABL SH2 DOMAINS. THE HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOPHYAL GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.

CC -!- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM RESPONSE ELEMENT (SRE) (BY SIMILARITY).

CC -!- DOMAIN: THE SH3 DOMAIN IS NECESSARY FOR THE LOCALIZATION OF THE PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-RICH REGION OF FOCAL ADHESION KINASE 1.

CC -!- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE YDYLVL MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -!- SIMILARITY: BELONGS TO THE CAS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U48853; AAA93381.1; -

DR EMBL; U28151; AAA93248.1; -

DR HSSP; P07751; 1BK2.

DR MGD; MGI:108091; Crkas.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00452; SH3DOMAIN.

DR PRODOM; PD000066; SH3; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS00002; SH3; 1.

KW Phosphorylation; SH3 domain; SH3-binding; Cell adhesion;

KW Alternative splicing.

FT DOMAIN 6 26

FT DOMAIN 74 87

FT DOMAIN 119 420

FT DOMAIN 426 618

FT SITE 639 647

FT DOMAIN 750 800

FT VARSPIC 1 4

SQ SEQUENCE 874 AA; 94256 MW; 5B9EDD76532BDBA CRC64;

Query Match 8.2%; Score 82; DB 1; Length 874;

Best Local Similarity 24.4%; Pred No. 9.4;

Matches 43; Conservative 17; Mismatches 60; Indels 56; Gaps 9;

Qy 4 VIAAARATQADAMAYATEKGRLLFFAGLREMYFG-----RALP 43

Db 583 LVACSRAPVEDAKQLASFLHGNASLLFRRTKAPGPGEGSSSLHPNPTDKASSIQSRPLP 642

Qy 44 SDANTMTKMSYSGIKG-YEPQDGAW---YNY-HTTGKGVLEKQDPKSFQVAVENILDLF 98

Db 643 SPKRFT---SODSPGQVSESGWMEDYDVHVGKKEFEK-----TQKELLERG 690

Qy 99 RTKKNYGAENGLHIAFLS-----NNDITGNSGSPVDFKNGLIGL 141

Db 691 NIMRQCK---GQLELQQLKQFERLEQVSRPIDHDLANWTAPQLVP--GRTGGL 740

RESULT 11

BCAL\_RAT STANDARD; PRT; 968 AA.

AC Q63767; Q63766;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE CRK-associated substrate (p130CAS) (Breast cancer anti-estrogen resistance 1 protein).

GN BCAR1 OR CRKAS OR CAS.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

[1]

RN SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RP TISSUE=Fibroblast;

RX MEDLINE=94349922; PubMed=8070403;

RA Sakai R., Iwamatsu A., Hirano N., Ogawa S., Tanaka T., Mano H., Yazaki Y., Hirai H.;

RT "A novel signalling molecule, p130, forms stable complexes in vivo with v-crk and v-src in a tyrosine phosphorylation-dependent manner.";

RL EMBO J. 13:3748-3756(1994).

[2]

TYROSINE PHOSPHORYLATION BY FOCAL ADHESION KINASE.  
MEDLINE=98030588; PubMed=9360983;  
Tachibana K., Urano T., Fujita H., Ohashi Y., Kamiguchi K., Iwata S.,  
Hirai H., Morimoto C.;  
"Tyrosine phosphorylation of Crk-associated substrates by focal  
adhesion kinase. A putative mechanism for the integrin-mediated  
tyrosine phosphorylation of Crk-associated substrates.";  
J. Biol. Chem. 272:29083-29090(1997).  
CC -!- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE  
FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION.  
CC IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY).  
CC -!- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1,  
CC ADAPTER PROTEIN-CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL.  
CC INTERACTS WITH NEPHROCYSTIN AND PTK2B (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: FOCAL ADHESIONS AND STRESS FIBERS.  
CC UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO  
CC THE MEMBRANE UPON TYROSINE PHOSPHORYLATION.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)  
CC AND A SHORT ISOFORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHER EXPRESSION IN LUNG,  
CC INTESTINE AND TESTIS.  
CC -!- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING  
CC MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN  
CC CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-  
CC BINDING SITES PUTATIVELY BIND CRK, NCK AND ABL SH2 DOMAINS. THE  
CC HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOPHYAL  
CC GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.  
CC -!- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM  
CC RESPONSE ELEMENT (SRE).  
CC -!- DOMAIN: THE SH3 DOMAIN IS NECESSARY FOR THE LOCALIZATION OF THE  
CC PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-RICH  
CC REGION OF FOCAL ADHESION KINASE 1.  
CC -!- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE  
CC YDVLH MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE  
CC PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE  
CC RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN  
CC MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX.  
CC -!- DISEASE: APPEARS TO HAVE A CENTRAL FUNCTION IN TRANSFORMATION OF  
CC SOME CELL TYPES.  
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE CAS FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; D29766; BAA06169.1; .  
DR EMBL; D29766; BAA06170.1; .  
DR HSSP; P07751; 1BK2  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF000018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PRODOM; PD000066; SH3; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW Phosphorylation; SH3 domain; SH3-binding; Cell adhesion;  
KW Alternative splicing.  
FT DOMAIN 97 159  
FT PRO-RICH.  
FT DOMAIN 168 181  
FT SUBSTRATE FOR KINASES.  
FT DOMAIN 213 514  
FT SER-RICH.  
FT SITE 733 741  
FT SH3-BINDING (POTENTIAL).  
FT DOMAIN 844 894  
FT DIVERGENT HELIX-LOOP-HELIX MOTIF.  
FT VARSPLIC 5 98  
FT MISSING (IN SHORT ISOFORM).  
SQ SEQUENCE 968 AA; 104262 MW; E861641BFD68D377 CRC64;  
-----  
Query Match 8.18; Score 81; DB 1; Length 968;  
Best Local Similarity 24.48; Pred. No. 13;  
Matches 43; Conservative 17; Mismatches 60; Indels 56; Gaps 9;

QY 4 VIAAARAIQADAMANAYAIKGRLLFFAGLRMPG-----RALP 43  
DB 677 LVACSRAPVEDAKQLASFLHGNASLLFRRTKAPGPGPEGSSSLHLNPTDKASSIQSRPLP 736  
QY 44 SDANTMTMSYGSIKG-YEQDQAW---YNY-HTTGKGVLEKQDKPSDEFAVOENILDF 98  
DB 737 SPKPT----SQSDPGQYENSEGGNMDYVHLQCKEFEK-----TKELLEKG 784  
QY 99 RPKNGRYAENCOLHIAFLS-----NNDITCGNSGSPVFDKNGRLIGL 141  
DB 785 NIVROGK---GQLEQLQKOFERLEQEVSRPIDHDLANWTPAQPLVP--GRTGGL 834  
RESULT 12  
MCRB\_METBA STANDARD; PRT; 433 AA.  
ID MCRB\_METBA  
AC P07955;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Methyl-coenzyme M reductase beta subunit (EC 1.8.-.-).  
GN MCRB.  
OS Methanosarcina barkeri.  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
OC Methanosarcinaceae; Methanosarcina.  
OX NCBI\_TaxID=2208;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fusaro / DSM 804;  
RX MEDLINE=87231011; PubMed=3502709;  
RA Bokranz M., Klein A.;  
RT "Nucleotide sequence of the methyl coenzyme M reductase gene cluster  
from Methanosarcina barkeri".  
RL Nucleic Acids Res. 15:4350-4351(1987).  
CC -!- FUNCTION: Reduction of methyl-coenzyme M (2-(methylthio)  
ethanesulfonic acid) with 7-mercaptoheptanoylthreosine phosphate  
to methane and an heterodisulfide.  
CC -!- CATALYTIC ACTIVITY: CH(3)-S-CoM + H-S-HTP = CH(4) + CoM-S-S-HTP.  
CC -!- COFACTOR: THE ENZYME COMPLEX BINDS TIGHTLY (BUT NOT COVALENTLY)  
CC TO TWO MOLECULES OF COENZYME F430. F430 IS A YELLOW NICKEL  
PORPHINOID.  
CC -!- PATHWAY: Methanogenesis; last step.  
CC -!- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO GAMMA CHAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Y00158; CAA68353.1; .  
DR PIR; A29525; A29525.  
DR HSSP; P11560; 1MRO.  
DR InterPro; IPR003179; MCR\_beta.  
DR Pfam; PF02241; MCR\_beta\_1.  
DR Pfam; PF02783; MCR\_beta\_N; 1.  
KW Methanogenesis; Oxidoreductase.  
FT INIT\_MET 0  
FT SEQUENCE 433 AA; 45289 MW; BABFAA43709361A9 CRC64;  
-----  
Query Match 8.08; Score 80.5; DB 1; Length 433;  
Best Local Similarity 20.18; Pred. No. 5.4;  
Matches 56; Conservative 34; Mismatches 74; Indels 115; Gaps 13;  
-----  
QY 2 KSVIAAARAIQADAMANAYAIKGRLLFFAGLRMPG--ALPSDANFTMRMSYG---- 55  
DB 40 RSVAVNLAIQIG-ALASGKMGKGQILGRGLNYDIVGNADAIAENVKLVQVDEGDDTN 98  
QY 56 ---SIKGYEPQDQAWYNYHTTGKGVLEKQDKPS-----DEFAVOENILDFR 99  
Matches 43; Conservative 17; Mismatches 60; Indels 56; Gaps 9;

```
DR PRINTS; PR00839; V8PROTEASE.  
DR PROSITE; PS00672; V8_HIS; 1.  
DR PROSITE; PS00673; V8_SER; 1.  
KW Hydrolase; Serine protease; Zymogen; Signal.  
FT SIGNAL      1    29  
FT PROPEP     30    68  
FT CHAIN      69   336  
FT ACT_SITE  119   119  
FT ACT_SITE  161   161  
FT ACT_SITE  237   237  
FT CONFLICT  109   109  
FT CONFLICT  125   125  
FT CONFLICT  145   145  
FT CONFLICT  193   193  
FT CONFLICT  229   229  
FT CONFLICT  259   261  
FT CONFLICT  268   270  
SQ SEQUENCE   336 AA; 36326 MW; 8B138D0C7996AA3E CRC64;  
  
Query Match      8.0%; Score 80; DB 1; Length 336;  
Best Local Similarity 31.5%; Pred.No.4.4;  
Matches 17; Conservative 13; Mismatches 24; Indels 0; Gaps 0;  
  
QY 123 TGGNSGSPVFDKNGRLIGLAPDGWNEAMSGDIIEPDQLORTISVDIFXVLFMID 176  
|||||::||::|::|::|::|::|::|::|::|:  
Db 233 TCGNSGSPVFKEKVETGIHGWGVNPFNGAFVNENVRNFLKNIEDIHFAND 286  
  
RESULT 14  
POLG_KUNJM STANDARD; PRT; 3433 AA.  
AC AC PI4335; Q82983;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein); Matrix  
DE Protein (Envelope protein M); Major envelope protein E; Nonstructural  
DE proteins NS1, NS2, NS4A and NS4B; Protease/helicase (EC 3.4.21.98)  
DE DE (NS3); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)].  
OS Kunjin virus (strain WPM61C)  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
CC Flavivirus.  
NCBI_TaxId=11078;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE-88089524; PubMed=2826659;  
RA Coia G., Parker M.D., Speight G., Byrne M.E., Westaway E.G.;  
RT "Nucleotide and complete amino acid sequences of Kunjin virus:  
RT definitive gene order and characteristics of the virus-specified  
R proteins";  
J. Gen. Virol. 69:1-21(1988).  
RL J. Gen. Virol. 69:1-21(1988).  
CC - FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE  
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.  
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.  
CC - CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the p6  
CC position, Cys or Thr in p1 and Ser or Ala in p1'.  
CC - CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {NH4}(N).  
CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
CC PROTEIN C AND MRNA.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to licensel@isb-sib.ch).  
-----  
DB EMBL: D00246; BAA00176.1;
```

```
DR PIR: A28697; GNRVKV.
DR HSP: P14336; ISVB.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR000069; Flavi_M.
DR InterPro: IPR001157; Flavi_NSI.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000336; Flavi_glycoprote.
DR InterPro: IPR001850; Flavi_helicase.
DR InterPro: IPR002535; Flavi_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00869; Flavi_glycoprot; 1.
DR Pfam: PF00948; Flavi_NSI; 1.
DR Pfam: PF00949; Flavi_helicase; 1.
DR Pfam: PF00972; Flavi_NS5; 1.
DR Pfam: PF01002; Flavi_NS2B; 1.
DR Pfam: PF01003; Flavi_capsid; 1.
DR Pfam: PF01004; Flavi_M; 1.
DR Pfam: PF01005; Flavi_NS2A; 1.
DR Pfam: PF01349; Flavi_NS4B; 1.
DR Pfam: PF01350; Flavi_NS4A; 1.
DR Pfam: PF01570; Flavi_propep; 1.
DR Pfam: PF01728; FtsJ; 1.
DR Pfam: PF02832; Flavi_glycop; 1.
DR ProDom: PD001496; Flavi_NSI; 1.
DR ProDom: PD001556; Flavi_glycoprote; 1.
DR SMART: SM00490; HELIC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
KW ATP-binding; Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 123
FT PROPEP 124 215
FT CHAIN 216 290
FT CHAIN 291 791
FT CHAIN 792 1143
FT CHAIN 1144 1374
FT CHAIN 1375 1505
FT CHAIN 1506 2124
FT CHAIN 2125 2273
FT CHAIN 2274 2528
FT CHAIN 2529 3433
FT DOMAIN 388 401
FT NP_BIND 1699 1706
FT SITE 1790 1793
FT DISULFID 293 320
FT DISULFID 350 406
FT DISULFID 364 395
FT DISULFID 382 411
FT DISULFID 480 578
FT DISULFID 595 626
FT CARBOHYD 138 138
FT CARBOHYD 921 921
FT CARBOHYD 966 966
FT CARBOHYD 998 998
SQ .SEQUENCE 3433 AA; 381363 MW; EB4B888A7D040B99 CRC64;

Query Match 8.0%; Score 80; DB 1; Length 3433;
Best Local Similarity 21.4%; Pred. No. 76;
Matches 43; Conservative 28; Mismatches 50; Indels 80; Gaps 10;

QY 52 MSYSGIKGYPQDGAWYN-----YHTGKGVLEKQDPKSGDEF--AVQENILDLFRTKNY 103
Db 1531 MTRGLLSYGQAGVMVGVGFTLWHTTKGAALMSGGRLLDPYWGSKEDRL-----CY 1584

PIR: A28697; GNRVKV.
DR HSP: P14336; ISVB.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR000069; Flavi_M.
DR InterPro: IPR001157; Flavi_NSI.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000336; Flavi_glycoprote.
DR InterPro: IPR001850; Flavi_helicase.
DR InterPro: IPR002535; Flavi_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00869; Flavi_glycoprot; 1.
DR Pfam: PF00948; Flavi_NSI; 1.
DR Pfam: PF00949; Flavi_helicase; 1.
DR Pfam: PF00972; Flavi_NS5; 1.
DR Pfam: PF01002; Flavi_NS2B; 1.
DR Pfam: PF01003; Flavi_capsid; 1.
DR Pfam: PF01004; Flavi_M; 1.
DR Pfam: PF01005; Flavi_NS2A; 1.
DR Pfam: PF01349; Flavi_NS4B; 1.
DR Pfam: PF01350; Flavi_NS4A; 1.
DR Pfam: PF01570; Flavi_propep; 1.
DR Pfam: PF01728; FtsJ; 1.
DR Pfam: PF02832; Flavi_glycop; 1.
DR ProDom: PD001496; Flavi_NSI; 1.
DR ProDom: PD001556; Flavi_glycoprote; 1.
DR SMART: SM00490; HELIC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
KW ATP-binding; Transmembrane; Nonstructural protein.
FT INIT_MET 1 1
FT CHAIN 1 123
FT PROPEP 124 215
FT CHAIN 216 290
FT CHAIN 291 791
FT CHAIN 792 1143
FT CHAIN 1144 1374
FT CHAIN 1375 1505
FT CHAIN 1506 2124
FT CHAIN 2125 2273
FT CHAIN 2274 2528
FT CHAIN 2529 3433
FT DOMAIN 388 401
FT NP_BIND 1699 1706
FT SITE 1790 1793
FT DISULFID 293 320
FT DISULFID 350 406
FT DISULFID 364 395
FT DISULFID 382 411
FT DISULFID 480 578
FT DISULFID 595 626
FT CARBOHYD 138 138
FT CARBOHYD 921 921
FT CARBOHYD 966 966
FT CARBOHYD 998 998
SQ .SEQUENCE 3433 AA; 381363 MW; EB4B888A7D040B99 CRC64;

Query Match 8.0%; Score 80; DB 1; Length 3433;
Best Local Similarity 21.4%; Pred. No. 76;
Matches 43; Conservative 28; Mismatches 50; Indels 80; Gaps 10;

QY 52 MSYSGIKGYPQDGAWYN-----YHTGKGVLEKQDPKSGDEF--AVQENILDLFRTKNY 103
Db 1531 MTRGLLSYGQAGVMVGVGFTLWHTTKGAALMSGGRLLDPYWGSKEDRL-----CY 1584
```

Search completed: May 25, 2003, 15:32:30  
Job time : 7.80731 secs

```
QY 104 G-----RYAENGQHLIAFL-----SNN-----DITGNSGSPV 131
Db 1385 GPPWKQLQHKWNGQDEVQMVVVEPKNVKNVQTKPGVKFTPEGEIGAVTLDFPTGTSGSPI 1644
QY 132 FDKNGRLIGLAFDG-----NWEAMSGDIE-----FEPDLQRTISVDI----- 168
Db 1645 VDKNGDVIGLYGNGVIMPNGSYISAIQGERMDEPVPAGFEPEMLRKQKITVLDLHPGAG 1704
QY 169 --RYVLEMDKMGOCPLIOE 187
Db 1705 KTRRIL-----POLIKE 1716

RESULT 15
Y030_UREPA
ID Y030_UREPA STANDARD; PRT; 747 AA.
AC Q9PRB5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein U0030.
GN U0030.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Letkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RL Nature 407:757-762(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002102; AAF30435.1;
DR Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7
FT POTENTIAL
SQ .SEQUENCE 747 AA; 85860 MW; 68FFD940C28D02AF CRC64;

Query Match 7.9%; Score 79.5; DB 1; Length 747;
Best Local Similarity 23.4%; Pred. No. 13;
Matches 39; Conservative 21; Mismatches 40; Indels 67; Gaps 8;

QY 27 RLFFAGLRREMPGRALPSDA---NFTWRMSYGS-IKGYEPQDGAWYNYHTGKGVLEKQD 82
Db 585 QLFSGG---YPGDVNPNSAIVSWRGSKYGSLQIAF----- 618
QY 83 PKSEFAVQENILDLFRTK---NYGRVAENGOLHIAFLSN-----NDITGNSG 128
Db 619 ---DREIKNESILDYGYGPKINNIIDGYQKVGEGYLNKLFNVTGRTVITSDEIGDLGSGSSG 675
QY 129 SPVFDKNGRLIGLAFD-----GNW-----EAMSGDIE 155
Db 676 SMITDSNFNVLVGIHFASLNSRAYGAPNDSMIGNLFAVQSDLSGDID 722

Search completed: May 25, 2003, 15:32:30
Job time : 7.80731 secs
```